

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 16:36:28 ; Search time 3028.25 Seconds
(without alignments)
11538.356 Million cell updates/sec

Title: US-09-868-474-1

Sequence score: 1 ccgtgaagatgcgcctcc.....aaaaaaaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pt:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_om:*
20: em_or:*
21: em_ov:*
22: em_pat:*
23: em_ph:*
24: em_pl:*
25: em_ro:*
26: em_sts:*
27: em_sy:*
28: em_un:*
29: em_vl:*
30: em_htgo_hum:*
31: em_htgo_inv:*
32: em_htgo_rod:*
33: em_htg_hum:*
34: em_htg_inv:*
35: em_htg_rod:*
36: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

8

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2118	100.0	2118	9	AF195651 Homo sapi
2	2098	99.1	2143	9	AK000639 Homo sapi
3	1380.4	65.2	1399	9	BC000395 Homo sapi
4	1322.8	62.5	1637	9	AL050286 Homo sapi
5	1056.6	49.9	180268	2	AC023648 Homo sapi
6	1056.6	49.9	202333	2	AC087884 Homo sapi
7	868.8	41.0	2447	10	AF287293 Homo sapi
8	462	21.8	473	9	HS000016 Homo sapi
9	352	16.6	2335	9	AB055368 Macaca fa
10	258.4	12.2	274	6	AB079372 Sequence
11	197.8	9.3	180268	2	AC023648 Homo sapi
12	197.8	9.3	336508	2	AC011600 Homo sapi
13	152.8	7.2	336508	2	AC011600 Homo sapi
14	152.6	7.2	212282	2	AC023070 Mus muscu
15	60.4	2.9	2157	9	AK026630 Homo sapi
16	57.6	2.7	1813	8	AF229811 Cucumis s
17	57.4	2.7	2164	3	DD1ACT15 Homo sapi
18	57.2	2.7	682	9	AK000414 Homo sapi
19	57	2.7	1649	8	AT074450 Arabidops
20	56.2	2.7	1912	9	AB050421 Macaca fa
21	56	2.6	6108	12	ASPDAXHY Artificial
22	56	2.6	6134	12	ASPDAXHC Artificial
23	56	2.6	6146	12	AF269236 Cloning v
24	56	2.6	6872	12	AF269235 Cloning v
25	56	2.6	10452	12	AF269238 Cloning v
26	56	2.6	11166	12	AF269237 Cloning v
27	55.8	2.6	9934	3	DD025660 Dictyostell
28	55.8	2.6	79594	9	AL138805 Human DNA
29	55.6	2.6	216410	2	AC027321 Homo sapi
30	55.4	2.6	3436	9	BC006201 Homo sapi
31	55.2	2.6	2488	9	AK027188 Homo sapi
32	55	2.6	384	6	AX185690 Sequence
33	55	2.6	1579	9	AF182416 Homo sapi
34	55	2.6	1896	5	AF024336 Homo sapi
35	55	2.6	3336	8	AF118843 Lycopersi
36	55	2.6	6105	12	ASPDAX3H Artificial
37	55	2.6	7218	12	ASPDAX3C Artificial
38	55	2.6	7218	6	I66494 Sequence 14
39	54.6	2.6	1316	9	AK027106 Homo sapi
40	54.6	2.6	3138	6	A21625 expression
41	54.4	2.6	754	9	HS040693 Homo sapi
42	54.2	2.6	1953	9	AB056420 Macaca fa
43	54.2	2.6	227203	2	AC011184 Homo sapi
44	54	2.5	4898	3	DM007629 Drosophila
45	53.6	2.5	1451	3	AF201906 Drosophila

ALIGNMENTS

RESULT 1
AF195651 2118 bp mRNA PRT 14-APR-2001
LOCUS Homo sapiens cervical cancer 1 protooncogene protein p40 mRNA,
DEFINITION complete cds.
ACCESSION AF195651
VERSION AF195651.1 GI:13624097
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2118)
AUTHORS Kim,J.W.
TITLE Identification of a new oncogene in human cancers
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2118)
AUTHORS Kim,J.W.
TITLE Direct Submission
JOURNAL Submitted (17-OCT-1999) Dept. of Obstet. & Gynecol., Kangnam St.

Db	2078	TTAAACCAGTAAGACTGCTTGAACCTGTGATGAGAAAAA	2137
Yy	2101	aaaaa 2106 	
Db	2138	AAAAA 2143	
RESULT	3		
LOCUS	BC000395		
DEFINITION	BC000395 1399 bp mRNA	PRI	12-JUL-2001
ACCESSION	BC000395		
VERSION	BC000395.1		
KEYWORDS	MGc.		
SOURCE	Human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1399)		
TITLE	Strausberg, R.		
JOURNAL	Direct Submission Submitted (15-NOV-2000) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA		
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov		
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: DCTP/DRP cDNA Library Preparation: Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL) DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC), Gaithersburg, Maryland; Web site: http://www.nisc.nih.gov/ Contact: nisc.mgc@nih.gov Shevchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S., Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R., Lim, M., Maduro, Q., V., Masello, C., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Snyder, B., Stanlipop, S., Thomas, P.J., Tjongson, E.E., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Zhang, L., H. and Green, E.D.		
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.llnl.gov Series: IRAL Plate: 1 Row: 9 Column: 18 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 4886472.		
SOURCE	Location/Qualifiers 1. 1399		
	/organism="Homo sapiens" /db_xref="taxon:9606 /db_xref="taxon:9606 /clone="MGC:8483 IMAGE:2821851" /tissue_type="Lung, small cell carcinoma" /clone_lib="NIH MGC-7" /lab_host="DH10B-R" /note="Vector: pOTB7" 97. 399 /codon_start=1 /product="DKFZP586A011 protein" /protein_id="AAH00395.1" /db_xref="GI:12653253" /translation="MLTSLPPLLRHRIKHTTIVHQDLAKLAKIGCOLTAQOEYK SACVLRGINSHGDECRKTWDEWQIJCISLKEAEKSLDLNHYVLLSTNLTGR"		
CDS			
BASE COUNT	334 a 324 c 335 g 406 t		
ORIGIN			
Query Match	65.2%	Score 1380.4	DB 9; Length 1399;
Best Local Similarity	99.6%	Prod No 0	

Matches 1384; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

QY 702 atcttgctctgaagagatgtttcttcaaccatctcttgagcatgaacacactcagct 761
DB 10 ATCTTGCTCTGAGAGAGTGTCTCTAACCATCTCTGAGCATGAACCACTCCAGCT 69
QY 762 ttgcaatgaagagcttgagccgggacatgtcttcaacttactctctctccttg 821
DB 70 TTGCACTGAAGAGCTTGAGCCGGGCAATGCTTCAACATCTTACCTGCTCTCCCTTG 129
QY 822 ttggaacatctgttgaagactatacaactgtgatcaaccaactgagcaagcttgca 881
DB 130 TTGGAACATCTGTTGAAGCTCATCACTCACTGATTTCACCACTGGAAGCTTTGGCA 189
QY 882 aagctggagatggagcagctgactctcaggaagtaaaatcgagctgtatctcgtgac 941
DB 190 AAGCTGGAGATGGCCAGCTGACTGCTCAGGAAGTAATTCGGCTTGTATCTCCGTGGC 249
QY 942 ctgaatctacgcatatgttggaagataaggtgtcgaacttgcttggaagatgctgag 1001
DB 250 CTGAATCTACGATATTTGTAAGATAGTGTGCAACTTGCTGGAGAAATGGCTGCAG 309
QY 1002 attctctgagcctgaagaaagctgactgtctctcttgctgacacagtgctctgc 1061
DB 310 ATTTCTCTGAGCTGAAAGAAAGCTGAGCTGTCTCTCTGTCGACAACTGCTCTCTGCTC 369
QY 1062 tccaccaactactctggagcaagcgctgaagcaactgaacacatgagagatgctgctg 1121
DB 370 TCCACCACACTACCTTGCGGCAAGGCGCTGAATGAACATGAGCGAGATGGCTTCTCG 429
QY 1122 caatgctatagatagcagtgaggaagcaaacacagcacttgcaagaagctgtgtgac 1181
DB 430 CAGCTGTATAGTATGAGCACTGAGCAAGCAACAGCACTTGCCAGCAAGTGTGTGTAC 489
QY 1182 tgttaagtgctgtggaagcagagaagaagagcaggggcataggtcttcaagcatgca 1241
DB 490 TGTTAAGTGTGTGGAAGGAGAGAGAGGCGAGGCGCATGCGCTTCAACAGATGGCAC 549
QY 1242 cctgtggagacgagacatctctctcagctagaactgaacaaacccctcttgag 1301
DB 550 CCTGTGGAGACGACACATCTCTCTCAGCTTAATACTGAACAAACCTCTTCTTACG 609
QY 1302 ggtggtcgtgtgaggtgcatcctgccccctcaataacttaataagctggaactgga 1361
DB 610 GGTGTGCTGAGAGTGCATCTGCTCCCTCAATAATTAATTAATTAATTAATTAATTA 669
QY 1362 gaagcctctacgagcttactgtgatgtgtcagctcagctcagtaagtaagctt 1421
DB 670 GCAGCTCTACTGAGGCTTTACTGTGATGTGTGATGATGATGATGATGATGATGATG 729
QY 1422 tgcacaggtggaatcccttacttgcttagagcagcatccacttccatgtaactatc 1481
DB 730 TGCACAGGTGGGANTCCTTATTTGGCTTAGAGCATGATCCATCTTACTTACTTACT 789
QY 1482 tgtgggtttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1541
DB 790 TGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 849
QY 1542 ctggagacatgtgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatgcatg 1601
DB 850 CTGGAGACATGTGTGTCATGATGATGATGATGATGATGATGATGATGATGATGATG 909
QY 1602 caaaaatcaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1661
DB 910 CAAAATCAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 969
QY 1662 tcaaacctctgaacaccttctctgataaggtgaacttgcaaggaatatctatgtgaag 1721
DB 970 TCAAACCTCTGAACACCTTTTCTGATGGGTAACTTGAGGAATATTTCTATTGGAAAG 1029
QY 1722 ataaacaggaagaaagtgctctctgacccctctctctctctctctctctctctctct 1781
DB 1030 ATAAACAGGAAGTAAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1089

```

```

QY 1782 atgtcttcttggtggtgtatatacagccctctgtgatacttcaactcgtgctccact 1841
DB 1090 ATTGTCTTTTGTGGGCTGTATTAACGCCCTGTGGAATCTTCAACTCTGCTGCCACT 1149
QY 1842 gtgaltgagcagtcacatgtaactgagtgagtgctctctctctggtgcatgatacac 1901
DB 1150 GTGATGCAAGCAGTCCAACTGTAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1209
QY 1902 ctgtgaagtgataatctctccagcctggggagatcaaggaaggtctgcatgttgta 1961
DB 1210 CTGTAAAGTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1269
QY 1962 agttgggttggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2021
DB 1270 AGTTGGGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1329
QY 2022 aatgcatgaagcactgttttaaacccaagttaaagactgcttgaaactgtgtatgaa 2081
DB 1330 AATGATGAACACTGTTTAAACCAAGTAAGACTGCTGAACCTGAAACCTGAAACCTG 1389
QY 2082 aaaaaaaaaa 2091
DB 1390 AAAAAAAAAA 1399

```

RESULT 4

HSMB00205 1637 bp mRNA PRI 18-FEB-2000

LOCUS Homo sapiens mRNA; cDNA DKFP586A011 (from clone DKFP586A011);

DEFINITION partial cds.

ACCESSION AL050286

VERSION AL050286.1 GI:4886472

KEYWORDS human.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 1637)

AUTHORS Koehler, K., Beyer, A., Mewes, H.W., Gassenhuber, J. and Wiemann, S.

TITLE Direct Submission

JOURNAL Submitted (15-MAY-1999) MIPS, Am Kioflerspitze 18a, D-82152

COMMENT Martinsried, GERMANY

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

sequenced by BMFZ (Biomedical Research Center at the Charité,

Berlin/Germany) within the cDNA Sequencing consortium of the German

Genome Project.

This clone (DKFP586A011) is available at the RZPD in Berlin.

Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059

Berlin-Charlottenburg, GERMANY; Email: clone@rzd.de Further

information about the clone and the sequencing project is available

at <http://www.mips.biochem.mpg.de/proj/cDNA/>.

FEATURES

1. 1637

Location/Qualifiers

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="DKFP586A011"

/clone_1lb="586 (synonym: hute1). Vector pSport1; host

DH10B; sites NotI + SalI/MluI"

/dev_stage="adult"

/tissue_type="uterus"

331..633

/gene="DKFP586A011"

331..633

/gene="DKFP586A011"

331..633

/note="DKFP586A011"

/note="similarity to D.melanogaster anon-66db"

/codon_start=1

/product="hypothetical protein"

/protein_id="CA43387.1"

/db_xref="GI:4886473"

/db_xref="SPTREMBL:O913X3"

/translation="MLTSTLPPPLLRHRLKTHITVHQDLAKLAKLGIGLTAQEVK

polya_site 1620
 BASE COUNT 392 a 378 c 382 g 485 t
 ORIGIN

Query Match 62.5%; Score 1322.8; DB 9; Length 1637;
 Best Local Similarity 99.8%; Pred. No. 0;
 Matches 1324; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 770 gaagccttgagccggccatgctctccacatcttaccgtccctccctctgttgagaca 829
 DB 312 GAAAGCCCTTGAGCCGGCCATGCTTCCACATCTTACCTCCCTCCCTGTTGAGACA 371
 QY 830 tgcgttaagaacatacaactgfatcccaactggaagaagccttgccaagcctgag 889
 DB 372 TCGTTTGAAGACTCATACACTGTGATTCACCACTGACAAAGCCTTGGCAAAAGCTGGG 431
 QY 890 gattggccagctgagctgctcagaagaataaactcggtctgtatctccgttgccgaatc 949
 DB 432 GATTGGCCAGCTGACTGCTCAGAGAAATGCGCTTGTATCTCCGTGCGCTGAATTC 491
 QY 950 taagcatattggtgaaataatggtgctgaactggtctggggaatggtcagaattctctg 1009
 DB 492 TACCATATTGGTGAAGATAGGTGTCAACTTGGCTGGAGAAATGGCTGCAGATTTCTCTG 551
 QY 1010 caagcctgaagaagcctgagctgctctctgctgacacagctggtccctgctccacaa 1069
 DB 552 CAGCCTGAAGAAGACTGAGCTGTCTCTTGTCTGCAACAGCTGCTGCTGCTCCACCAA 611
 QY 1070 ctaccttgggaacaagcgtctgaatgaacccalggagcggatgcatgtctcctgaatcgt 1129
 DB 612 CTACCTTGGGACAGCGCTGAATGACCAATGAGCGGATGCGATGCTCTGCAAGTCGTA 671
 QY 1130 taagtatggcagtgaggaagaacaaacagcactctgccaagaagctctgtgtgaactgta 1189
 DB 672 TAGTATGACAGTGCAGAACAAACAGCACTTCCAGCAAGTCTGTGTGATCTGTTAAGT 731
 QY 1190 gctgtagagcagagagagagagagcagggccatgggtctcacaagatgagcacactgtg 1249
 DB 732 GTGTGAGAGGACAGAGAGAGAGAGCGGCGCATGGCTTCAACAGATGGCAACCTGTGGG 791
 QY 1250 aactgacacatctcctcctcagctagaaactgaacaaacccctctgtctaggggtgtcc 1309
 DB 792 AACTGCGAGCAATCTCTCTCAAGCTAGAACGTAACGTAACAAACCTCTTGGCTAGGGGTGCTC 851
 QY 1310 gctgtagagctcaatcctcctcctcctcctcctcctcctcctcctcctcctcctcctc 1369
 DB 852 GTGTGAGAGTGTCACTGCTCCCTCATATATTAATAGCTGGAACCTGCGACAGCCTC 911
 QY 1370 tactgggcttctactgtagtctgctcagctcagctcctaggaagcagcctcttgccccag 1429
 DB 912 TACGGGCTTTTACTGTGATGTGTTCAGTTCATGTCTTAGAAGTCAAGCTTTGGCCCCAG 971
 QY 1430 gctggaactcctatttggcttaagactgataccacttcacatgtaacttaactcctgtggt 1489
 DB 972 GTGGGAATCCCTTATTTGGCTTAGGACTGATCCATTCCATGACTTACATCTGTGGGT 1031
 QY 1490 ttgttttctgtttagaanaatttttggctgtgtggaagaagcagcctcctgtggtgagca 1549
 DB 1032 TTTGTTTGTGTGTGAANAATTTGTGGCTGTGAAGAACAGCACTCTTGTGGCTGGAGCA 1091
 QY 1550 ctctgctcagctgactgactggtgttccctccacatcctctgatataccacaaatc 1609
 DB 1092 CTGTGTCCGCGCATGTACTTGGGTGTTCCTCCATCCCTTCTGTGATATACCAAAATC 1151
 QY 1610 aagctgttctgttttttgcacactcagcgtgagcattgggtctaacactccttttcaacc 1669
 DB 1152 AAGTTGTTTGTGTTTGTGACCTTCACTGAGGCAATGAGGCTTCTTTTCAAAACC 1211
 QY 1670 tctggaacactttctgtagggtaactgctcagaataatctctatgtggaagaagtaacag 1729
 DB 1212 TCTGAACACCTTTTCTGATGGGTAACTGACAGAAATATTTCTATGTGAAAGATTAACAG 1271

QY 1730 aagtaacagtgctctctgagccctccctccatglttctcagcctcactcctcactgctc 1789
 DB 1272 AAGTACAACTGCTTCTTGAGCCCTTCCATCATGTCTTACGCTTACCTCATGTTCTT 1331
 QY 1790 tctctggtctgatatcaagcctctctgagctcctcactcactcctcactcactgagca 1849
 DB 1332 TTTCTGGGCTGTATTAACAGCCCTCTGTGATCTTCAACCTGTGCTGCTCCACATGTATGCA 1391
 QY 1850 gcaagtcacactgtaactgacagtgagctgctcctcctctgagccatggaatcaacactgta 1909
 DB 1392 GCAAGTCAACTGATTAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1451
 QY 1910 tactaactcctccagcctctgagagatcaagaagagctgcatagtaagtaagttggt 1969
 DB 1452 TACTATTACTGCTCCAGCTGCGGAGATCAGAGAGAGTCTGATGTTAGTAAGTTGGGT 1511
 QY 1970 ttagcttctgtgtgcatcaagtgacttagagctctgttaataactattgttaatgcatg 2029
 DB 1512 TTAGCTTTTGTGTGTCATCACTGACTTGAAGTCTGTAATTAATTAATTAATTAATGATG 1571
 QY 2030 aagcactgttcttaaccccaagtaagactgcttgaacactgttgatgaaaaaataa 2089
 DB 1572 AAGCACTGTTTAAACCAACTAAAGACTGCTTGAACCTGTTATGATGAAAAAATAA 1631
 QY 2090 aaaaaa 2095
 DB 1632 AAAAAA 1637

RESULT 5
 AC023648/C
 LOCUS
 DEFINITION
 Homo sapiens chromosome 12 clone RP11-624H4 map 12, WORKING DRAFT
 ACCESSION
 AC023648
 VERSION
 AC023648.3 GI:784355
 KEYWORDS
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 SOURCE
 human.
 ORGANISM
 Homo sapiens

REFERENCE
 1 (bases 1 to 180268)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 12, clone RP11-624H4
 Unpublished
 JOURNAL
 TITLE
 AUTHORS

REFERENCE
 2 (bases 1 to 180268)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Beda, F., Boguslavsky, L.,
 Boukhalter, B., Brown, A., Burkett, G., Campiano, A., Castle, A.,
 Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
 Deatellano, K., Dewar, K., Dodge, S., Domingo, M., Doyle, M.,
 Feneclor, J., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hago, B., Heaford, A., Horton, L.,
 Howland, J., Iliov, I., Johnson, R., Jones, C., Kahn, L., Karatas, A.,
 Klein, J., Landers, T., Laroque, R., Lehoczy, J., Levine, R.,
 Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M.,
 McEwan, P., McGurk, A., McKernan, K., McPheters, R., Meldrum, J.,
 Menus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Naylor, J.,
 Norman, C. H., O'Connell, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,
 Peterson, K., Piere, R., Pisan, C., Pollara, V., Raymond, C.,
 Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,
 Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talmas, J., Testfay, S., Theodore, J., Tirrell, A.,
 Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.
 Direct Submission
 Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On May 17, 2000 this sequence version replaced gi:7650699.
 All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/repeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L6516
 Center clone name: 624_H_4
 ----- Summary Statistics
 Sequencing vector: M13; M77815; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 153867 bases at least Q40
 Consensus quality: 164187 bases at least Q30
 Consensus quality: 170599 bases at least Q20
 Insert size: 176000; agarose-fp
 Insert size: 176868; sum-of-contigs
 Quality coverage: 5.3 in Q20 bases; agarose-fp
 Quality coverage: 5.3 in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 35 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 *
 1 1101: contig of 1101 bp in length
 * 1102 1201: gap of 100 bp
 * 1102 1201: contig of 1057 bp in length
 * 2259 2358: gap of 100 bp
 * 2359 2358: contig of 1103 bp in length
 * 3462 3561: gap of 100 bp
 * 3562 3561: contig of 1078 bp in length
 * 4640 4739: gap of 100 bp
 * 4740 4739: contig of 1691 bp in length
 * 6431 6530: gap of 100 bp
 * 6531 6530: contig of 1124 bp in length
 * 7655 7754: gap of 100 bp
 * 7755 7754: contig of 1616 bp in length
 * 9371 9470: gap of 100 bp
 * 9471 9470: contig of 1017 bp in length
 * 10487 10487: contig of 1017 bp in length
 * 10488 10587: gap of 100 bp
 * 10588 11847: contig of 1260 bp in length
 * 11848 11947: gap of 100 bp
 * 11948 13526: contig of 1579 bp in length
 * 13527 13626: gap of 100 bp
 * 13627 15225: contig of 1599 bp in length
 * 15226 15325: gap of 100 bp
 * 15326 17019: contig of 1694 bp in length
 * 17020 17119: gap of 100 bp
 * 17120 18225: contig of 1106 bp in length
 * 18226 18325: gap of 100 bp
 * 18326 20585: contig of 2260 bp in length
 * 20586 20685: gap of 100 bp
 * 20686 22352: contig of 1667 bp in length
 * 22353 22452: gap of 100 bp
 * 22453 26283: contig of 3831 bp in length
 * 26284 26383: gap of 100 bp
 * 26384 28561: contig of 2178 bp in length
 * 28562 28661: gap of 100 bp
 * 28662 31275: contig of 2614 bp in length
 * 31276 31375: gap of 100 bp
 * 31376 34050: contig of 2675 bp in length
 * 34051 34150: gap of 100 bp
 * 34151 37375: contig of 3225 bp in length
 * 37376 37475: gap of 100 bp
 * 37476 40705: contig of 3230 bp in length
 * 40706 40805: gap of 100 bp
 * 40806 44888: contig of 4083 bp in length

FEATURES
 source
 * 44889 44988: gap of 100 bp
 * 44989 51688: contig of 6700 bp in length
 * 51689 51788: gap of 100 bp
 * 51789 60033: contig of 8245 bp in length
 * 60034 60133: gap of 100 bp
 * 60134 66633: contig of 6500 bp in length
 * 66634 66733: gap of 100 bp
 * 66734 72372: contig of 5639 bp in length
 * 72373 72472: gap of 100 bp
 * 72473 80057: contig of 7585 bp in length
 * 80058 80157: gap of 100 bp
 * 80158 89044: contig of 8887 bp in length
 * 89045 89144: gap of 100 bp
 * 89145 99209: contig of 10065 bp in length
 * 99210 99309: gap of 100 bp
 * 99310 109103: contig of 9800 bp in length
 * 109110 109209: gap of 100 bp
 * 109210 118651: contig of 9442 bp in length
 * 118652 118751: gap of 100 bp
 * 118752 130023: contig of 11272 bp in length
 * 130024 130123: gap of 100 bp
 * 130124 140909: contig of 10786 bp in length
 * 140910 141009: gap of 100 bp
 * 141010 154860: contig of 13851 bp in length
 * 154861 154960: gap of 100 bp
 * 154961 180268: contig of 25308 bp in length.
 Location/Qualifiers
 1..180268
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="12"
 /map="12"
 /clone.lib="RP11-624H4"
 /clone.lib="RP11-624H4" Human Male BAC"
 1..1101
 /note="assembly-fragment"
 1202..2258
 /note="assembly-fragment"
 2359..3461
 /note="assembly-fragment"
 3562..4639
 /note="assembly-fragment"
 4740..6430
 /note="assembly-fragment"
 6531..7654
 /note="assembly-fragment"
 7755..9370
 /note="assembly-fragment"
 9471..10487
 /note="assembly-fragment"
 10588..11847
 /note="assembly-fragment"
 11948..13526
 /note="assembly-fragment"
 13627..15225
 /note="assembly-fragment"
 15326..17019
 /note="assembly-fragment"
 17120..18225
 /note="assembly-fragment"
 18326..20585
 /note="assembly-fragment"
 20686..22352
 /note="assembly-fragment"
 22453..26283
 /note="assembly-fragment"
 26384..28561
 /note="assembly-fragment"
 28662..31275
 /note="assembly-fragment"
 31376..34050
 /note="assembly-fragment"
 34151..37375
 /note="assembly-fragment"
 37476..40705
 /note="assembly-fragment"
 40806..44888
 /note="assembly-fragment"

QY	1979	ggtgtgcatgaagtaagctatgaagctctgttaaaccttcttbaaagctgaagacagt	2038
Db	70211	GTGTGTGATCATGCTACTTACAGTTCTGTATACTTTTGTAAATGCATGACGACTGT	70270
QY	2039	ttttaaaccacaagtaaaagcttgcctgttaaacctgttcatgtgaagaa	2081
Db	70271	TTTAAACCCAAAGTAAGACTCTCTTGAACACCTGTTGATGAGAA	70313
RESULT	7		
LOCUS	AF287293	2447 bp	02-AUG-2001
DEFINITION	Mus musculus cervical cancer receptor (MCCR) mRNA, complete cds.		
ACCESSION	AF287293		
VERSION	AF287293.1	GI:15077021	
KEYWORDS			
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 2447)		
TITLE	Hwang,J.H., Kim,J.W. and Hwang,S.Y.		
JOURNAL	Mouse cervical cancer receptor		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 2447)		
TITLE	Hwang,J.H., Kim,J.W. and Hwang,S.Y.		
JOURNAL	Direct Submission		
FEATURES	Submitted (13-JUL-2000) Biochemistry, Hanyang University, 1271 Sa-dong, Ansan, Kyunggi-do 425-791, Korea		
source	1. 2447		
gene	/organism="Mus musculus"		
CDS	/strain="C3H/He"		
	/db_xref="taxon:10090"		
	/cell_line="p19"		
	/cell_type="embryonic carcinoma"		
	1. 2447		
	/gene="MCCR"		
	31. 1113		
	/gene="MCCR"		
	/codon_start=1		
	/product="cervical cancer receptor"		
	/protein_id="AAK83032.1"		
	/db_xref="GI:15077022"		
	/translation="MALSRVCAKAAALMGSTVAFPPVYTRLDGRSGPMARPRSK LHSKPADKVNLSIVYTKRINALGSHRELGREPPFVALYTEMGIDMLVADKKR ARRIKADMMKONLKFHOLSYREMEHLQRFDRIDICLAFVGLSTIPFANLIVFLMYL FPRLLKHEFTPEKQIDFLVDVYHGLRRSHSEVYTHLRASFTSVSHKLRDLTDLC KSYVSGHPAAODVLAIRDCSTYPLGFSLOLAQOMALSOAMLLPYLDPLLRRL KSHYTHHOLDRALAKIKIGOLITPQVFKSACYLRLGLNSTHADRCAAMLGEMHLHISC SLKEPELSTLHNHNVLSINTLETR"		
BASE COUNT	535 a 683 c 599 g 630 t		
ORIGIN			
Query Match	41.0%: Score 868.8; DB 10; Length 2447;		
Best Local Similarity	81.6%: Pred. No. 1.9e-22;		
Matches 1057; Conservative	0; Mismatches 227; Indels 12; Gaps 4;		
Db	1	ctgttaagaatgagcttcacagagtgctgagctcggtcgctgtgtgagagctgagag	60
	23	CTGTGAAGATGGCGCTGTCCAGGGGTGCTGGGCTGGGCTGTGTGGGGTTGGACTG	82
QY	61	tcacccctggaacatttctaccccggaagctggaactgtgtcgctgtgactgtgag	120
Db	83	TGCGCCCTGGACCCCTTTGTACACCGGAGCGTGCACCTGGCCGCTGGGCTGGTGGC	142
QY	121	gggcccctcggtctctcaagcttcaaccttctccaaagcgagatgtgaagaactgtat	180
Db	143	GAGCCCTCGTGGCTTCAAGCTTACCTTTCTCCGAAGCGGACGTGAAGAACTGATTT	202
QY	181	cttatgtgttaaccacaagcaaaagatataatggagaaataccatggttcttggatg	240

Db	203	CTTAGGTTGGTACCAACAAAGACGATTAAAGCATTTACAGCATCTGTAACATGCTTTCTTGGGGCGCTC	262
OY	241	attcccccgccttcatalaccctgtacaacaactcctcgaagaagatgtcagatgltatgg	300
Db	263	ATTTCGCCCGCTCTTCATGGCTTTATACACAACCTTCATGAAAGAAATACAGATGTTATGG	322
OY	301	ctgagtcgcaaaaagctcagaagaataaagaacaatctctgagaagacaaataaagaatttc	360
Db	323	CTGATGGCAAAAAGCTAGAGAAGATTAAGGCGACATGTGGAAACAAAACCTTAAAGTTTC	382
OY	361	atcaaatcccaataccggagatbtagagcattttagaacagtctccgcaagaagtcacacagt	420
Db	383	ATCAACACTCTTACCGGGAATAGAGACTTTTGAAGACAGTTCCGACGAGATATACACCAGT	442
OY	421	gctcttcctcagglattattcattccattccaccttctgcgaactacactgblctctgtcaa	480
Db	443	GTCCTTCCTGTTGGGCTCATCTATCCATTCACCCCTTTGCCAACTACCTGGCTTTTGTGCTAA	502
OY	481	tgtaacctgtttcccaagcaactactcgtataagagatttctggagccccaagaacaaactg	540
Db	503	TGTACCTGTTTCTTAAGGCAACTCTGTAGTCAAGGACTTTTGTGGACCCCAACAACTAAATGG	562
OY	541	attccttagatatcacaagcttccgcgaagaagcagctcccaaccaaaatattatattt	600
Db	563	ATTTCCTTGGACGTTTATCATGATGCTCTCGAAGSGGGTCCCACTCGGAAGTCATCACACTC	622
OY	601	tagaaaaagtcaltccctcctcattcctgtatgcagagactccggtgtgcctgtcacagatcgt	660
Db	623	TAAAGAAAGGCGCAGCACTTGTGTTTCTCATGACAAATCTCCGAGGACGCTGACAGACTGT	682
OY	661	gcacaaagataaagcgctgagaccacccaagaataatagatatctctgtctcgtagaaggt	720
Db	683	GCACCAAGGTGCAGAGTGGGCAACCCACGACGACACAGAGACGTCTGCTCTCAAGATT	742
OY	721	gttctcttaacaactccctcttggcgtatgaacaacaactccagagcttgcagctggaagccttga	780
Db	743	GCTTCTTACTACTACCCCTTGGGCTTACGCAAGTCCAAACCTTCGACATGAGAGAGCTTGA	802
OY	781	gcgcggccatgctcttcacactcttaccctgcctccctctgttctgtagaatcgttgaaga	840
Db	803	GCCAGGCAATGCTCTACACACTTACTCTGCCCCCGCTTGGTGAAGGACGCGCTTAAAGA	862
OY	841	ctcgtaaacgttgatctaccacactctgaaagaagcttggcaagcttgggaatttggcagc	900
Db	863	GTCACACCAACCGTGATCACCACAGCTGGACAGGGCTCTGGCAAAAGCTGGGATTGGCCAGC	922
OY	901	tgaacttctcaggaagtaaaaaactgcgcttgttattatccgttgccttgatattcctacgtat	960
Db	923	TGACTGCAACGAAAGTAAATGTGGCTTGGCATCTTGCTGGCTTGAACCTACCCACATTG	982
OY	961	gtgaagaatagtgctcgaacttggctctggagaaatgctgcagattcctctgcagcctgaaag	1020
Db	983	CTWAGCAGCAGGTGTGCGAGCCGTGGGGGAATGGTTGCACATCTCTGTACCCCTGAAG	1042
OY	1021	aagctgagctgtctctctctgtctgcacaaogtggctctgtcttcacacaactacacttggga	1080
Db	1043	AACCCGAGCTGTCCCTCTTCTCTCAACAAGCTGGTCTCGTCTCCACCAACTACCTTGAGA	1102
OY	1081	caagagcctaatatgaa--ccatlggaacgaatgcaatcttgcctgcgaagctcgtatagatagc	1138
Db	1103	CAAGGCGCTGATGTAACCCCAAGGAAAGACGCGAATTGCTCTGCAAGTCACTAAGTATGT	1162
OY	1139	agtgacgaacaacaacagcacttgcacagaaagtctgtgtgactgttlaagtgltg99ag	1198
Db	1163	GGTGTTGGGACCAAGAAAGCACCTGTGCAGC--TGTCTGTGGGCACACATA-----GTGTG	1213
OY	1199	gcagagaagagagcaggggccatgtggctctcaacgaatgycacacactgttgggaactgcga	1258
Db	1214	GGAGGCAAGCGGAGTGGGTTTCACGGGGCTTTCAGAGCAAGCTTACCCCTCGGGAGGCTATAGA	1273
OY	1259	cattcctctcaacagcttagaacttgtaaaacaacctct	1294
Db	1274	CATTCTCT-TACACGCTTGAGCTCTTACAAACCACTGT	1308

Qy 1897 cacacctgttaaggtactaattacttcgccagccttgggagatcaggagaagtctgcatagt 1956
|||||

Query Match	16.68;	Score 352;	DB 9;	length 2335;
Best Local Similarity	95.28;	Pred. No. 4e-83;		
Matches 396;	Conservative 0;	Mismatches 15;	Indels 5;	Gaps 3

```

Qy 1707 attcattggaagatacaggaagtagctcttgcacccctcctcaatgtctc 1766
Db 1863 ATTCTATGGAAGATACAGGAAGTACAGTGGTCTTCTTACCCCTCTCTCAATGTTC 1922
Qy 1767 tagccttaacctcctcatgtcttcttctggtctgtatacagccctcgttgatctca-- 1824
Db 1923 TAGCCTTACCTCTCATGTGTCTTTCCTGCTGATTAACA-CCCCCTGTGATTCAGC 1981
Qy 1825 actctgctcctcactgtatgtacagcagctgaactgaactgaactgtgctctc 1884
Db 1982 GCTTTGCTGCTCCACCTGTATGACAGCAATCACTGTAACCTGACAGTGGCTCTTCTT 2041
Qy 1885 tgggccaatgatacacaactgtgaagtaactaactcctcctggtggtgagatcagga 1944
Db 2042 TGGGCAATGATCAACACTGTAAGTACTAATATTTGCGCCAGCTGGGAGATCAGAGAGA 2101
Qy 1945 ggtctgact--agttagaagatgtggtttagcttctgtgtgtatcaagtagactagatc 2002
Db 2102 GGTCTGCAATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2161
Qy 2003 tctgtaataacttattgtgaatgcatgaagcactgtctttaaaccagaataaagactgct 2062
Db 2162 TCTGTAATAACGTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2221
Qy 2063 tgaagacctgtgtatggaagaaagaaagaaagaaagaaagaaagaaagaaagaa 2118
Db 2222 TGAAACCTGTGTAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2277

RESULT 10
AX079372 274 bp DNA PAT 22-FEB-2001
LOCUS Sequence 116 from Patent WO0107611.
DEFINITION AX079372
ACCESSION AX079372
VERSION AX079372.1 GI:13158932
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 274)
AUTHORS Baker,K.P., Goddard,A. and Wood,W.I.
TITLE Human polypeptides and methods for the use thereof
JOURNAL Patent: WO 0107611-A 116 01-FEB-2001;
Genentech, Inc. (US)
PUBMED 11607611
SOURCE location/Qualifiers
1..274
/organism="Homo sapiens"
/db_xref="taxon:9606"

BASE COUNT 79 a 53 c 52 g 84 t 6 others
ORIGIN

```

```

Query Match 12.28; Score 258.4; DB 6; Length 274;
Best Local Similarity 95.68; Pred. No. 3.3e-58;
Matches 262; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

Qy 238 gtcattcccccgtctcattatctcctgtacacaaatctcagaagaagatgcagatgtatc 297
Db 1 GTCATTTCCCGCGCTTTATATCTGTAACACAAATTTTCAAGAAAGATTGACAGATGTTAT 60
Qy 298 gggctgatacgaagagtagaagaataaagacaaatatlgtggaagacacataaagt 357
Db 61 GGGCTGAGGCCAAAGAGGTAGAGATTAAGACAAATATGTGAGACACACATATAAGT 120
Qy 358 ttctcaactcctacccggagatgagacatttgaaacagcttcgcccaagaagtcacca 417
Db 121 TTNAATCAANTTCATACCGGAGATGAGACATTTGAGACAGTTCCGCCAAGANTGCACA 180
Qy 418 agtctcttccagtagatattcctcactcacttctgccaactaccatgctgtctgc 477
Db 181 AGTGTNTTCTCTAGTATATTTTCATTCACACTTTTGCCCAANTACTGCTGTTTTTTTGC 240

```

```

Qy 478 taatgacctgttcccaagcactactatcag 511
Db 241 TAAATGTAACCTGTTTCCAGGCAATATATGATCAG 274

```

```

RESULT 11
AC023648
LOCUS Homo sapiens chromosome 12 clone RP11-624H4 map 12, WORKING DRAFT
DEFINITION SEQUENCE, 35 unordered pieces.
AC023648
VERSION AC023648.3 GI:7884355
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 180268)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
JOURNAL Homo sapiens chromosome 12, clone RP11-624H4
PUBMED 1160268
TITLE Unpublished
JOURNAL 2 (bases 1 to 180268)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E.,
Anderson,S., Baldwin,J., Barna,N., Beda,E., Boguslavsky,L.,
Bouhaghalter,B., Brown,A., Burkett,G., Campoliano,A., Castle,A.,
Choehel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
Deatellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
Fenster,J., Ferreira,P., Fitzhugh,W., Forrest,C., Gage,D.,
Galagan,J., Gardyna,S., Glinde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heathford,A., Horton,L.,
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., Landers,T., Laroque,K., Lechoczy,J., Levine,R.,
Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N., McCarthy,M.,
McEwan,P., McKernan,K., McPheters,R., Meldrum,J.,
Meneses,L., Mihova,T., Miranda,C., Mieng,V., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,T.M.,
Peterson,K., Pierre,N., Pisanic,C., Pollara,V., Raymond,C.,
Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
Severly,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Subramanian,A., Talamas,J., Testifye,S., Theodore,J., Tirrell,A.,
Travers,M., Triggillo,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
Zody,M.
PUBMED 1160268

```

```

TITLE Direct Submission
JOURNAL Submitted (16-FEB-2000) Whitehead Institute/MIT Center for Genome
RESEARCH, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT All repeats were identified using RepeatMasker:
http://ftp.genome.washington.edu/RM/RepeatMasker.html

```

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 16516

Center clone name: 624.H4

Summary Statistics

Sequencing vector: M13; M77815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 153867 bases at least 940

Consensus quality: 164187 bases at least 930

Consensus quality: 170599 bases at least 920

Insert size: 176000; agarose-fp

Quality coverage: 5.3 in Q20 bases; agarose-fp

Quality coverage: 5.3 in Q20 bases; sum-of-ctrls

* NOTE: This is a 'working draft' sequence. It currently consists of 35 contigs. The true order of the pieces

Jackson, L. E., Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudan, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korah, J., Kovar, C., Krawiec, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L., Li, J., Li, Z., Lichte, O., Ilean, C., Liu, C., Liu, J., Liu, W., Lounsges, H., Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Meshkini, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawlawi, E., McLeod, M. P., Meador, M., Mel, G., Melker, M., Miner, G., Moser, M., Neal, D., Newton, J., Newton, K., Morgan, M., Morris, S., Moser, M., Neale, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokoko, S., Ogutu, K., Okunuga, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pichens, R., Primus, E., Pu, L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherer, S., Scott, G., Shen, H., Shoshitari, N., Stone, I., Sodergren, E., Sonalkar, T., Sparks, A., Stanley, H., Stinson, H., Sutton, A., Svatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H., Tansari, J., Taylor, C., Taylor, T., Tellro, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, J., Vera, V., Villalón, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K., Wu, C., Williams, G., Williamson, A., Wleczek, R., Wooden, S., Weinstock, G., and Gibbs, R.	REFERENCE
TITLE	JOURNAL
AUTHORS	REFERENCE
TITLE	JOURNAL
COMMENT	

```

Chemistry: Dye-primers Body: 12% of reads
Chemistry: Dye-terminator Big Dye: 88% of reads
Assembly program: Phrap; version 0.990339
Consensus quality: 318894 bases at least Q30
Consensus quality: 342195 bases at least Q40
Consensus quality: 354034 bases at least Q20
Estimated insert size: 350560; sum-of-configs estimation
Quality coverage: 0x in Q20 bases; agarose-1p estimation
Quality coverage: 3.4x in Q20 bases; sum-of-configs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It is uncorrected.

```

```

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_drift\_data.html)
* NOTE: This sequence may represent more than one clone.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 49 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 34038: contig of 34038 bp in length
* * 34039 34138: gap of unknown length
* * 34139 34289: contig of 29151 bp in length
* * 34289 63390: gap of unknown length
* * 63390 63399: gap of unknown length
* * 63399 81254: contig of 17865 bp in length
* * 81255 81354: gap of unknown length
* * 81355 97046: contig of 15692 bp in length
* * 97047 97146: gap of unknown length
* * 97147 109167: contig of 12021 bp in length
* * 109168 109267: gap of unknown length

```


Db 304434 GAAGTAAATCGG 304422

RESULT 13

AC011600 LOCUS AC011600 336508 bp DNA HTG 27-JUN-2001
DEFINITION Homo sapiens chromosome 12 clone RP11-392116, WORKING DRAFT
SEQUENCE, 49 unordered pieces.
AC011600
AC011600 28 GI:14547729
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 336508)
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaralunge,H.C., Are,J.R., Banks,T., Barbarella,J.,
Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Caron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy,Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
Hollins,B., Homsli,F., Howard,S., Huber,J., Hulik,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudan,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Loulsegod,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Meshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,M., Mawhney,E., McLeod,M.P., Meador,M.,
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenko,S.,
Ogutu,M., Okunolu,G., Oreguene,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Plimus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M.,
Ruiz,S., Saverly,G., Scherer,S., Scott,G., Shen,H., Shooshbari,N.,
Stinson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wlaczek,R., Wooden,S.,
Worley,K., Wu,C., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.
Unpublished
Direct Submission
2 (bases 1 to 336508)
Worley,K.C.
Submitted (08-OCT-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 25, 2001 this sequence version replaced gi:14190620.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMLU
Center clone name: RP11-392116

Summary Statistics

Sequencing vector: Plasmid; M7789
Sequencing vector: M13; 108821
Chemistry: Dye-terminator Big Dye 88% of reads
Chemistry: Dye-terminator Big Dye 88% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 318894 bases at least Q40
Consensus quality: 342195 bases at least Q30
Consensus quality: 354034 bases at least Q20
Estimated insert size: 350560, sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-1p estimation
Quality coverage: 3.4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)

* NOTE: This sequence may represent more than one clone.

* NOTE: This is a 'working draft' sequence. It currently

* consists of 49 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1
34038: contig of 34038 bp in length

* 34039
34138: gap of unknown length

* 34139
63289: contig of 29151 bp in length

* 63290
63389: gap of unknown length

* 63390
81254: contig of 17865 bp in length

* 81255
81354: gap of unknown length

* 81355
97046: contig of 15692 bp in length

* 97047
97146: gap of unknown length

* 97147
109167: contig of 10201 bp in length

* 109168
109267: gap of unknown length

* 109268
123513: contig of 14646 bp in length

* 123914
124013: gap of unknown length

* 124014
133886: contig of 9873 bp in length

* 133887
133986: gap of unknown length

* 133987
144693: contig of 10707 bp in length

* 144694
144793: gap of unknown length

* 144794
154244: contig of 9451 bp in length

* 154245
154344: gap of unknown length

* 154345
163322: contig of 8978 bp in length

* 163323
163422: gap of unknown length

* 163423
172826: contig of 9404 bp in length

* 172827
172926: gap of unknown length

* 172927
180526: contig of 7600 bp in length

* 180527
180626: gap of unknown length

* 180627
188940: contig of 8314 bp in length

* 188941
189040: gap of unknown length

* 189041
195336: contig of 6596 bp in length

* 195337
195736: gap of unknown length

* 195737
201841: contig of 6105 bp in length

* 201842
201941: gap of unknown length

* 201942
209588: contig of 7647 bp in length

* 209589
209688: gap of unknown length

* 209689
217437: contig of 7749 bp in length

* 217438
217537: gap of unknown length

* 217538
222495: contig of 4958 bp in length

* 222496
222595: gap of unknown length

* 222596
227782: contig of 5187 bp in length

* 227783
227882: gap of unknown length

* 227883
233115: contig of 5233 bp in length

* 233116
233215: gap of unknown length

* 233216
238189: contig of 4974 bp in length

* 238190
238289: gap of unknown length

* 238290
244865: contig of 6576 bp in length

* 244866
244965: gap of unknown length

* 244966
249984: contig of 5019 bp in length

* 249985
250084: gap of unknown length

* 250085
254445: contig of 4361 bp in length

* 254446
254545: gap of unknown length

* 254546
259133: contig of 4588 bp in length


```

* 259134 259233: gap of unknown length
* 259234 264486: contig of 5253 bp in length
* 264487 264586: gap of unknown length
* 264587 267850: contig of 3264 bp in length
* 267851 267950: gap of unknown length
* 267951 271465: contig of 3515 bp in length
* 271466 271565: gap of unknown length
* 271566 274968: contig of 3403 bp in length
* 274969 275068: gap of unknown length
* 275069 278811: contig of 3743 bp in length
* 278812 278911: gap of unknown length
* 278912 281845: contig of 2934 bp in length
* 281846 281945: gap of unknown length
* 281946 285805: contig of 3860 bp in length
* 285806 285905: gap of unknown length
* 285906 289213: contig of 3308 bp in length
* 289214 289313: gap of unknown length
* 289314 293460: contig of 4147 bp in length
* 293461 293560: gap of unknown length
* 293561 297333: contig of 3773 bp in length
* 297334 297433: gap of unknown length
* 297434 300059: contig of 2626 bp in length
* 300060 300159: gap of unknown length
* 300160 302533: contig of 2374 bp in length
* 302534 302633: gap of unknown length
* 302634 306196: contig of 3563 bp in length
* 306197 306296: gap of unknown length
* 306297 309165: contig of 2869 bp in length
* 309166 309265: gap of unknown length
* 309266 311791: contig of 2526 bp in length
* 311792 311891: gap of unknown length
* 311892 315049: contig of 3158 bp in length
* 315050 315149: gap of unknown length
* 315150 317792: contig of 2643 bp in length
* 317793 317892: gap of unknown length
* 317893 320763: contig of 2871 bp in length
* 320764 320863: gap of unknown length
* 320864 323189: contig of 2326 bp in length
* 323190 323289: gap of unknown length
* 323290 325971: contig of 2682 bp in length
* 325972 326071: gap of unknown length
* 326072 32641: gap of 2570 bp in length
* 328641 328741: gap of unknown length
* 328742 331416: contig of 2675 bp in length
* 331417 331516: gap of unknown length
* 331517 334019: contig of 2503 bp in length
* 334020 334119: gap of unknown length
* 334120 336508: contig of 2389 bp in length.

```

FEATURES

```

source
1..336508
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="12"
  /clone="RP11-392116"
BASE COUNT 104868 a 63616 c 63734 g 99395 t 4895 others
ORIGIN

```

```

Query Match      7.2% Score 152.8: DB 2: Length 336508;
Best Local Similarity 98.7% Pred. No. 1.8e-29;
Matches 154; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 128 tcggcttcaaacgtccaccttcccaagcagatgtgaagaacttgatgtctatgt 187
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235359 TAGGTCTTCAAGCTTCACCTTTCTCCAAAGAGATGTGAAGACTTATGCTTATGT 235418

QY 188 ggtacccaagacaagcgattaaatgggaataaccatcgcttcttggtgcgtcattcc 247
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235419 GGTACCAAGACAAGCGATTAAATGGGAATATCCATCGTTCTTGCGTCGTCATTCC 235478

QY 248 ccgctctatctcgtgtacaacatctcatgaag 283
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 235479 CCGCTCTATGCTCTGTACACATCTTCATGAAGG 235514

```

```

RESULT 14
AC023070/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

AC023070 212282 bp DNA HTG 09-MAY-2001
Mus musculus clone RP23-11118, WORKING DRAFT SEQUENCE, 14 unordered
pieces.
AC023070
AC023070.5 GI:13940736
HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 212282)
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-11118
Unpublished
2 (bases 1 to 212282)
Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Baldwin, J., Barna, N., Beckert, R., Beda, F.,
Boguslavskiy, L., Boukhgalter, B., Brown, A., Burkett, G., Castle, A.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,
Deatellano, K., Dewar, K., Domingo, M., Doyle, M., Fenesior, J.,
Ferreira, P., FitzHugh, W., Forrest, C., Gage, D., Galagan, J.,
Gardyna, S., Grant, G., Hagos, B., Heaford, A., Horton, L.,
Howland, J., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J.,
Landers, T., Lehoczy, J., Levine, R., Lien, C., Liu, G., Locke, K.,
Macdonald, P., Margulis, N., McEwan, P., McGuck, A., McKernan, K.,
McHeeters, R., Meldrum, J., Menus, L., Morrow, J., Naylor, J.,
Norman, C. H., O'Connor, T., O'Donnell, P., Olivar, T. M., Peterson, K.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rothman, D.,
Roy, A., Santos, R., Severy, P., Spencer, B., Stange, Thoman, N.,
Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Tirelli, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wymann, D., Ye, W. J.,
Zimmer, A. and Zody, M.
Direct Submission
Submitted (08-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On May 4, 2001 this sequence version replaced g1:13123357.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIDR
Web site: http://www-seg.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L5130
Center clone name: 11.L118
----- Summary Statistics
Sequencing vector: M13; M77815; 37% of reads
Sequencing vector: Plasmid; n/a; 63% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 208323 bases at least Q40
Consensus quality: 208672 bases at least Q30
Consensus quality: 210447 bases at least Q20
Insert size: 198000; agarose-fp
Insert size: 210982; sum-of-coverage
Quality coverage: 8.4 in Q20 bases; agarose-fp
Quality coverage: 7.9 in Q20 ba.
NOTE: This is a 'working draft' sequence. It currently
consists of 14 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 2135: contig of 2135 bp in length
* 2136 2235: gap of 100 bp

```



```

/db xref="taxon:9606"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone="KAT11312"
/clone_id="KAT"
/note="cloning vector pME18FL3"
261. 962

```

CDS

BASE COUNT ORIGIN	736 a	368 c	450 g	603 t
----------------------	-------	-------	-------	-------

11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525
 526
 527
 528
 529
 530
 531
 532
 533

Query Match	2.9%	Score 60.4	DB 9	Length 2157
Best Local Similarity	71.8%	Pred. No. 4.4e-05		
Matches	79	Conservative	0	Mismatches 31
				Indels 0
				Gaps 0

[illegible]

Search completed: December 20, 2001, 19:18:59
Job time: 9751 sec



GenCofe version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 17:28:29 ; Search time 205.75 Seconds
(without alignments)
8825.346 Million cell updates/sec

Title: US-09-868-474-1
2118
Sequence: 1 ctgtgaagatgcgcctcc.....aaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 930621 seqs, 428662619 residues
Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_1101.*
1: /SID88/gcgdata/geneseq/NA1980.DAT.*
2: /SID88/gcgdata/geneseq/NA1981.DAT.*
3: /SID88/gcgdata/geneseq/NA1982.DAT.*
4: /SID88/gcgdata/geneseq/NA1983.DAT.*
5: /SID88/gcgdata/geneseq/NA1984.DAT.*
6: /SID88/gcgdata/geneseq/NA1985.DAT.*
7: /SID88/gcgdata/geneseq/NA1986.DAT.*
8: /SID88/gcgdata/geneseq/NA1987.DAT.*
9: /SID88/gcgdata/geneseq/NA1988.DAT.*
10: /SID88/gcgdata/geneseq/NA1989.DAT.*
11: /SID88/gcgdata/geneseq/NA1990.DAT.*
12: /SID88/gcgdata/geneseq/NA1991.DAT.*
13: /SID88/gcgdata/geneseq/NA1992.DAT.*
14: /SID88/gcgdata/geneseq/NA1993.DAT.*
15: /SID88/gcgdata/geneseq/NA1994.DAT.*
16: /SID88/gcgdata/geneseq/NA1995.DAT.*
17: /SID88/gcgdata/geneseq/NA1996.DAT.*
18: /SID88/gcgdata/geneseq/NA1997.DAT.*
19: /SID88/gcgdata/geneseq/NA1998.DAT.*
20: /SID88/gcgdata/geneseq/NA1999.DAT.*
21: /SID88/gcgdata/geneseq/NA2000.DAT.*
22: /SID88/gcgdata/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2118	100.0	2118	AAF30809	Human cervical can
2	1817.4	85.8	2002	AAH45403	Human cDNA encodin
3	1312.8	62.0	1645	AAH18148	Lung cancer associ
4	1095.6	51.7	1197	AAH30152	Human secreted pro
5	741	35.0	783	AAH03869	Human cDNA clone (
6	509.6	24.1	583	AAH03930	Human cDNA clone (
7	384.8	18.2	422	AAZ42807	Human 5' EST isola
8	282.6	13.3	381	AAO04230	Human secreted pro
9	282.4	13.3	387	AAO04231	Human secreted pro
10	258.4	12.2	274	AAF93295	Lung carcinoma cDN
11	189.2	8.9	398	AAH30637	Human colon cancer

C	12	100.6	4.7	458	21	AAA43579	Human secreted exp
	13	87	4.1	936	22	AAE58252	Oligonucleotide D1
	14	87	4.1	936	22	AAE58254	Oligonucleotide D1
	15	87	4.1	936	22	AAE58257	Oligonucleotide D1
	16	87	4.1	936	22	AAE58259	Oligonucleotide D2
	17	87	4.1	936	22	AAE58262	Oligonucleotide D2
	18	87	4.1	936	22	AAE58265	Oligonucleotide D1
	19	84.6	4.0	936	22	AAE58255	Oligonucleotide D1
	20	84.6	4.0	936	22	AAE58254	Oligonucleotide D1
	21	84.6	4.0	936	22	AAE58257	Oligonucleotide D1
	22	84.6	4.0	936	22	AAE58259	Oligonucleotide D2
	23	84.6	4.0	936	22	AAE58262	Oligonucleotide D2
	24	84.6	4.0	936	22	AAE58265	Oligonucleotide D1
	25	57.6	2.7	1813	21	AAA64012	DNA encoding a 9-h
	26	57.6	2.7	1813	21	AAZ49782	Cucumber hydropero
	27	55.8	2.6	1954	21	AAA3441	CDNA encoding huma
	28	55.8	2.6	3288	22	AAE05613	Human secreted pro
	29	55.8	2.6	3393	22	AAE05613	Human secreted pro
	30	55	2.6	384	22	AAH70111	Human cervical can
	31	55	2.6	1895	21	AAA38747	zebrafish Hsp-bind
	32	54.6	2.6	282	21	AAE08972	Fusarium venenatum
	33	54.6	2.6	900	21	AAH71625	Murine aspartate p
	34	54.6	2.6	3138	12	AAQ11712	Shuttle vector PMU
	35	53.6	2.5	995	21	AAE69107	Human secreted pro
	36	53.4	2.5	2329	21	AAE79959	Human secreted pro
	37	53.2	2.5	331	21	AAE98501	Human colon cancer
	38	53.2	2.5	1820	20	AAZ00434	Human secreted pro
	39	53	2.5	1531	21	AAZ297040	Human secreted pro
	40	53	2.5	1623	17	AAH14925	Human secreted pro
	41	53	2.5	1623	19	AAH64062	I cell replating f
	42	53	2.5	2013	22	AAH34803	Plasmid pSP6K-mTfR
	43	53	2.5	2138	21	AAE59412	Human colon cancer
	44	52.8	2.5	1067	22	AAE33104	Human secreted pro
	45	52.4	2.5	1480	22	AAE07771	Human secreted pro

ALIGNMENTS

RESULT 1	AAF30809	standard; cDNA: 2118 BP.
ID	AAF30809	
XX	AAF30809;	
AC	AAF30809;	
XX	21-JUN-2001 (first entry)	
DT	21-JUN-2001 (first entry)	
XX	Human cervical cancer 1 protooncogene.	
DE	Human cervical cancer 1 protooncogene.	
XX	Cervical cancer 1 protooncogene; HCCR-1; oncogene; human;	
KW	Lung cancer; leukemia; lymphoma; kidney cancer; liver cancer;	
KW	ovarian cancer; diagnosis; gene therapy; ss.	
XX		
OS	Homo sapiens.	
XX		
FT	Key	Location/Qualifiers
FT	CDS	9..1091
FT	sig_peptide	9..83
FT	mat_peptide	84..1088
FT	polyA_signal	2008..2012
FT		/*tag- d
XX	WO200127149-A1.	
XX	19-APR-2001.	
XX	30-MAR-2000; 2000WO-KR00284.	
XX	15-OCT-1999; 59KR-0044811.	
XX		

PA (KIMJ/) KIM J W.
 XX KIM JW;
 XX WPI: 2001-290710/30.
 DR P-PSDB: AAB20492.
 XX
 PT Novel human cervical cancer 1 protooncogene is useful in the diagnosis
 PT of various cancers, e.g., leukemia, lymphoma, kidney, liver, lung,
 PT ovary, and uterine cervix cancers -
 XX
 PS Claim 1; Page 56-59; 67pp: English.
 XX
 CC The present sequence is that of novel human cervical cancer 1
 CC protooncogene (HCCR-1) cDNA. Differential display RT-PCR analysis
 CC provided a 206 bp fragment, CC214, that was expressed in cervical
 CC cancer, metastatic tissue and CUMC-6 cervical cancer cells but not
 CC in healthy tissue. CC214 was used as a probe to screen a phage
 CC lambda gII human lung embryonic fibroblast cDNA library,
 CC identifying the present full-length HCCR-1 cDNA clone. The HCCR-1
 CC protooncogene encodes a 360-amino acid protein (see AAB20492) that
 CC is markedly hydrophobic and possesses a characteristic single
 CC membrane-spanning domain and pre-secretory signal peptide. The
 CC protooncogene is overexpressed in cervical cancer tissues, cervical
 CC cancer cell lines and lung cancer cell lines. Overexpression is
 CC also observed in leukemia, lymphoma, kidney, liver and ovarian
 CC cancers. The protooncogene is used in claimed kits for the
 CC diagnosis of cancer. Antisense sequences are used in a claimed
 CC process for treating or preventing cancer in humans. Also claimed
 CC are a vector comprising the protooncogene, a microorganism
 CC transformed with the vector, especially Escherichia coli JM109/HCCR-1
 CC (KCTC 0667BP), and a process for preparing the encoded protein.
 XX
 SO Sequence 2118 BP; 530 A; 496 C; 491 G; 601 T; 0 other;

Query Match 100.0%; Score 2118; DB 22; Length 2118;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2118; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ctggaagatgagcgtcccaagggtgctggtggtcgtgctggtggtggtggtcgcag 60
 DB 1 ctggaagatgagcgtcccaagggtgctggtggtcgtggtggtggtggtcgcag 60
 QY 61 tcaccccgagcatttctcaccgagagctgcaactgtgctgctggtcgtggtggtggt 120
 DB 61 tcaccccgagcatttctcaccgagagctgcaactgtgctgctggtcgtggtggtggt 120
 QY 121 gggccctcggtcttcaaaagcttcaccttctccaaaggcagatgtgaagaactgtat 180
 DB 121 gggccctcggtcttcaaaagcttcaccttctccaaaggcagatgtgaagaactgtat 180
 QY 181 ctatgtgtgtaacaaagaaagcgttaattgtggaataacacgtttcttgggtgtc 240
 DB 181 ctatgtgtgtaacaaagaaagcgttaattgtggaataacacgtttcttgggtgtc 240
 QY 241 attcccccgtctctatactctgtacacatcttcagaaagagatgagatgtatcgg 300
 DB 241 attcccccgtctctatactctgtacacatcttcagaaagagatgagatgtatcgg 300
 QY 301 ctgtagtccaaaagcgttagaagaaataagacaatatgtggaagacacataataagttc 360
 DB 301 ctgtagtccaaaagcgttagaagaaataagacaatatgtggaagacacataataagttc 360
 QY 361 atcaacttcacacggagagatgtgagcaattgagacagttccgcaagaagtcacaagt 420
 DB 361 atcaacttcacacggagagatgtgagcaattgagacagttccgcaagaagtcacaagt 420
 QY 421 gctcttctcaggtatattctcacttccactttgccaactactgcttcttggtaa 480
 DB 421 gctcttctcaggtatattctcacttccactttgccaactactgcttcttggtaa 480
 QY 481 tgaactgttccaggaactactgacagcatcttctgaccccaaaaacaactg 540

DB 481 tgaactgttccaggaactactgacagcatcttctgaccccaaaaacaactg 540
 QY 541 attcttagatatactatctgcttccggaagcagttcccaaaaataatattat 600
 DB 541 attcttagatatactatctgcttccggaagcagttcccaaaaataatattat 600
 QY 601 tagaaagtcacatccctccatcttctgagtcagagactccgggtggtcgaagtcgt 660
 DB 601 tagaaagtcacatccctccatcttctgagtcagagactccgggtggtcgaagtcgt 660
 QY 661 gcaaccaagatacagctggtgtaaccaccagcaatatacttctgctgtagaagt 720
 DB 661 gcaaccaagatacagctggtgtaaccaccagcaatatacttctgctgtagaagt 720
 QY 721 gttctctaacatctctctggtgcatgaaccactccagcttgcagtgaaagcttga 780
 DB 721 gttctctaacatctctctggtgcatgaaccactccagcttgcagtgaaagcttga 780
 QY 781 gccgggcatgtcttcacatcttaccctccctccctggtggaacatcgttgaaaga 840
 DB 781 gccgggcatgtcttcacatcttaccctccctccctggtggaacatcgttgaaaga 840
 QY 841 ctatacaactgtgattcacaaccactggaagagcttggcaagctggtggtcagc 900
 DB 841 ctatacaactgtgattcacaaccactggaagagcttggcaagctggtggtcagc 900
 QY 901 tgcctctgaagaagaatgaatgctgtgtatctccgtggtcgtggaatctagaatgt 960
 DB 901 tgcctctgaagaagaatgaatgctgtgtatctccgtggtcgtggaatctagaatgt 960
 QY 961 gtaagaatggtgtcgaactgtgctggtggaatgtgctgagattctcctgcagccgaag 1020
 DB 961 gtaagaatggtgtcgaactgtgctggtggaatgtgctgagattctcctgcagccgaag 1020
 QY 1021 aagctgagctgtctctcttctgtgacaaagctgtcctcctccaaactacattgga 1080
 DB 1021 aagctgagctgtctctcttctgtgacaaagctgtcctcctccaaactacattgga 1080
 QY 1081 caagcgctggaatgaacacatgtagcgagatggttctcctgaagtcgtatataagcag 1140
 DB 1081 caagcgctggaatgaacacatgtagcgagatggttctcctgaagtcgtatataagcag 1140
 QY 1141 tgcaggaacaacaacagcacttgcagcaagaagctgtgtgacgtttaagttgtggaagc 1200
 DB 1141 tgcaggaacaacaacagcacttgcagcaagaagctgtgtgacgtttaagttgtggaagc 1200
 QY 1201 agagagaagagcagggccatggtctcacaagcatggaacacctgtggaactcagaca 1260
 DB 1201 agagagaagagcagggccatggtctcacaagcatggaacacctgtggaactcagaca 1260
 QY 1261 ttccctcagcttagaacttgaaacaaacctctgtgctgaggggtggtcgtgtggaagtg 1320
 DB 1261 ttccctcagcttagaacttgaaacaaacctctgtgctgaggggtggtcgtgtggaagtg 1320
 QY 1321 catccgtcccccctcaataactactaaatagctggaactggaacagccttactggtgctt 1380
 DB 1321 catccgtcccccctcaataactactaaatagctggaactggaacagccttactggtgctt 1380
 QY 1381 tactgtgatgtgttcagttcatgtctcctaggaagtcagcttggcccaagtggaatcct 1440
 DB 1381 tactgtgatgtgttcagttcatgtctcctaggaagtcagcttggcccaagtggaatcct 1440
 QY 1441 tattgtgcttagaacctgacacacttccatgtaactactaactcgtgtgttctgtgtgc 1500
 DB 1441 tattgtgcttagaacctgacacacttccatgtaactactaactcgtgtgttctgtgtgc 1500
 QY 1501 tgttagaaaattttgtgcgtgtgaaacagcaactccttggctgtagagcactgtgtcat 1560
 DB 1501 tgttagaaaattttgtgcgtgtgaaacagcaactccttggctgtagagcactgtgtcat 1560
 QY 1561 gcatgtactgtggtgtttccctccatccttctgatatgacaaaataaagttgttgg 1620

Db	1561	gcagtagcttggggttccctccatcccttcttgcataatgacaaataaataagttgtttg	1620
Qy	1621	ttttttgtcaccttcaactgtgcagtgaggctaacaccacttcttlttcaaacctctgaacact	1680
Db	1621	tttttttcaaccttcaactgtgcagtgatbgttgtaacacacttcttlttcaaacctctgaacact	1680
Qy	1681	ttttctgatggtaacttgcaggaataattcatttgaaaagaatacaacgaaagtacaaagt	1740
Db	1681	ttttctgatgttgtaacttgcaggaataattcatttgaaaagaatacaacgaaagtacaaagt	1740
Qy	1741	ctctctgaaccccttccctcoaatbgtttctagcccttcaactcccatctgtcttcttctggcgtg	1800
Db	1741	ctctctgaaccccttccctcoaatbgtttctagcccttcaactcccatctgtcttcttctggcgtg	1800
Qy	1801	attcagagccctcttgatattcttaacactctgtcttccctcaacttgatgacagatccaact	1860
Db	1801	attcagagccctcttgatattcttaacactctgtcttccctcaacttgatgacagatccaact	1860
Qy	1861	gtaaactgaacagtgagctgcctctctctcttcggccatgagatcacacctgtaaagtacataact	1920
Db	1861	gtaaactgaacagtgagctgcctctctctcttcggccatgagatcacacctgtaaagtacataact	1920
Qy	1921	gcccaagcttggggagatcattgagagaggtctgcatagttaagtgttggtttagctttgt	1980
Db	1921	gcccaagcttggggagatcattgagagaggtctgcatagttaagtgttggtttagctttgt	1980
Qy	1981	gtgtgcattcaatgtaacttagagttctctgtaataactattgttaaatgcatgaagcactgtt	2040
Db	1981	gtgtgcattcaatgtaacttagagttctctgtaataactattgttaaatgcatgaagcactgtt	2040
Qy	2041	ttaaaccccaagtaaaagactgcttgaacacctgttgaatggaataaaaaaa	2100
Db	2041	ttaaaccccaagtaaaagactgcttgaacacctgttgaatggaataaaaaaa	2100
Qy	2101	aaaaaaaaaaaaaaaaaa 2118	
Db	2101	aaaaaaaaaaaaaaaaaa 2118	
RESULT	2		
AAH45403			
ID	AAH45403	standard; cDNA; 2002 BP.	
AC	AAH45403;		
XX			
XX			
DT	05-SEP-2001	(first entry)	
XX			
DE		Human cDNA encoding ribosomal protein L14.22.	
XX			
KW		Human; ribosomal protein L14.22; malignant neoplasm; haemopathy;	
KW		HIV infection; immunological disease; inflammatory disorder; cystostatic;	
KW		haemostatic; virucide; immunomodulatory; antiinflammatory; ss.	
XX			
OS		Homo sapiens.	
XX			
FH	Key	Location/Qualifiers	
FT	CDS	413..1024	
FT		/*tag= a	
FT		/product= "Ribosomal protein L14.22"	
XX			
PN	WO200138389-A1.		
PD			
XX	31-MAY-2001.		
XX			
PF	20-NOV-2000; 2000WO-CN00471.		
XX			
PR	24-NOV-1999; 99CN-0124098.		
XX			
PA	(BIOR-) BIOROAD GENE DEV LTD SHANGHAI.		
XX			
PI	Mao Y, Xie Y;		
XX			
DR	WPI; 2001-355906/37.		

DR	P-PSDB;AAG62443.
xx	
PI	Human ribosomal protein L14.22 and encoded polynucleotide, used in
PT	diagnosis and treatment of malignant tumors, hemopathy, human
PR	immunodeficiency virus infection, immunological diseases and
PI	inflammation -
xx	
PS	Claim 6; Page 19-20; 30pp; Chinese.
xx	
CC	This invention relates to human ribosomal protein L14.22 and cDNA
CC	encoding it. The invention includes a vector containing the cDNA, a host
CC	cell transformed with the vector and an antibody directed against human
CC	ribosomal protein L14.22. The cDNA and protein sequences are used in the
CC	diagnosis and treatment of malignant neoplasms, hemopathy, human
CC	immunodeficiency virus (HIV) infection, immunological diseases and
CC	various inflammatory disorders. Use of the protein and nucleotide
CC	sequence may result in cytostatic, haemostatic, virucide,
CC	immunomodulatory and anti-inflammatory activity. This sequence represents
CC	cDNA encoding human ribosomal protein L14.22.
xx	
SQ	Sequence 2002 BP; 456 A; 494 C; 467 G; 585 T; 0 other;
Query Match	85.8%; Score 1817.4; DB 22; Length 2002;
Best Local Similarity	94.1%; Pred. NO. 0;
Matches 1947; Conservative	0; Mismatches 6; Indels 116; Gaps
OY	1 ctgtgaagaatgcgctcccaaggagtgtagtcgtcgctgcggctgtgtggcctgcacg 60
Dd	
OY	50 ctgtgaagaatgcgctcccaaggagtgtagtcgtcgctgcggctgtgtggcctgcacg 109
OY	61 tcaccoccttgaacattttgtcacccggagcgtgcaacttgcgtctgtgcctgtgtgg 120
Dd	
OY	110 tcaccoccttgaacattttgtcacccggagcgtgcaacttgcgtctgtgcctgtgtgg 169
OY	121 gggccctcgtgcttccaagcttcaaccttcccacaaggcagaatgtaagaacttgatg 180
Dd	
OY	170 gggccctcgtgcttccaagcttcaaccttcccacaaggcagaatgtaagaacttgatg 229
OY	181 ctatagtgtlaaaccaagaacaaaagcatlaatggaataacatcgtttcttgggtcgtc 240
Dd	
OY	230 ctatagtgtlaaaccaagaacaaaagcatlaatggaataacatcgtttcttgggtcgtc 289
OY	241 attccccgcgttctatatctcgtgtacacaacttcaatgaagaatgtgcagargttatgg 300
Dd	
OY	290 attccccgcgttctatatctcgtgtacacaacttcaatgaagaaga----- 329
OY	301 ctgatgtccaaaaagcgttagaagaataagaacaatatgtgtggaagcaacaataaagtttc 360
Dd	
OY	330 ----- 329
OY	361 atcaactlcatlacoggagatgagacatttgagacagttccgccaaagcgtloaccaagt 420
Dd	
OY	330 ----- agttccgcaagaagcgtloaccaagt 353
OY	421 gtccttcttagtgtattattccatccacgttttggcaactgaacctgggtctcttgcta 480
Dd	
OY	354 gtccttcttagtgtattattccatccacgttttggcaactgaacctgggtctcttgcta 413
OY	481 tttacactgtttcccaaagcaactactgatatgaagcatttctggcccccaacaacaacgy 540
Dd	
OY	414 tggactcgtttcccaaagcaactactgatatgaagcatttctggcccccaacaacaacgy 473
OY	541 attctctaaratatctatcgtcttccggaaagcgtcccaaccagaatatattagttatt 600
Dd	
OY	474 attctctaaratatctatcgtcttccggaaagcgtcccaaccagaatatattagttatt 533
OY	tggaaagtcacaccttcacattcttcgtagcagaagatccggtgcgtctgcacagatctgt 660
Dd	
OY	534 tggaaagtcacaccttcacattcttcgtagcagaagatccggtgcgtctgcacagatctgt 593
OY	661 gcaaccaagatacagcgtgtgttacccaccagcaatatgatatcttggctctgagaagat 720
Dd	

Db 594 gcaaccaagatacagctggtgtaaccacagcaatataatgctgctgagagat 653
 QY 721 gttctcaacacatccttgggagcaacacacacagccttgcagctgaagccttga 780
 Db 654 gttctcaacacatccttgggagcaacacacacagccttgcagctgaagccttga 713
 QY 781 gccgggacatgcttcaacatcttaactgctcctcctcttggatagacatgcttga 840
 Db 714 gccgggacatgcttcaacatcttaactgctcctcctcttggatagacatgcttga 773
 QY 841 ctctaacacatgcttcaacatcttaactgctcctcctcttggatagacatgcttga 900
 Db 774 ctctaacacatgcttcaacatcttaactgctcctcctcttggatagacatgcttga 833
 QY 901 tgaactgctcaggaagtaaaatcagccttcttctcctgagcctgaatcttaacagatg 960
 Db 834 tgaactgctcaggaagtaaaatcagccttcttctcctgagcctgaatcttaacagatg 893
 QY 961 gtaagaataggtgtcgaactgtgctgaggaatgagctcagatctcctcagcctgaag 1020
 Db 894 gtaagaataggtgtcgaactgtgctgaggaatgagctcagatctcctcagcctgaag 953
 QY 1021 aagctgagctgtctctcttcttctgacaaagtgtcctcctcctcaacacatcttggga 1080
 Db 954 aagctgagctgtctctcttcttctgacaaagtgtcctcctcctcaacacatcttggga 1013
 QY 1081 caaggccttgaatgaacacatgagcagatgagctgtcctcctcctcctcaacacatcttggga 1140
 Db 1014 caaggccttgaatgaacacatgagcagatgagctgtcctcctcctcctcaacacatcttggga 1073
 QY 1141 tgcaggaacaaacagcactctgacagcaagctgtgtgtaactgttaagtgtgtgagagc 1200
 Db 1074 tgcaggaacaaacagcactctgacagcaagctgtgtgtaactgttaagtgtgtgagagc 1133
 QY 1201 agagaagagagcagagagcagctgagccttcaacagcctgtgagcagcctgtgagcagc 1260
 Db 1134 agagaagagagcagagagcagctgagccttcaacagcctgtgagcagcctgtgagcagc 1193
 QY 1261 ttcctcctcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1320
 Db 1194 ttcctcctcaacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1253
 QY 1321 catctgttccctcctcaataataataataataataataataataataataataataataata 1380
 Db 1254 catctgttccctcctcaataataataataataataataataataataataataataataata 1313
 QY 1381 taactgatagtgtcagctcaatcagctcaggaagatcagcttctgcccaggttggaatcct 1440
 Db 1314 taactgatagtgtcagctcaatcagctcaggaagatcagcttctgcccaggttggaatcct 1373
 QY 1441 taactgatagtgtcagctcaatcagctcaggaagatcagcttctgcccaggttggaatcct 1500
 Db 1374 taactgatagtgtcagctcaatcagctcaggaagatcagcttctgcccaggttggaatcct 1433
 QY 1501 tgttagaataattttgtgctgtgtaaaacagacactccttggctgagagcactgtgttcacat 1560
 Db 1434 tgttagaataattttgtgctgtgtaaaacagacactccttggctgagagcactgtgttcacat 1493
 QY 1561 gcatgatactgggtgttccctcctcaatccttctgataatgagcaaaatacaagtgttttg 1620
 Db 1494 gcatgatactgggtgttccctcctcaatccttctgataatgagcaaaatacaagtgttttg 1553
 QY 1621 tttttgtcaactcacttgcagtgatggtgtaaacacacttcttcaaacacacttgaacac 1680
 Db 1554 tttttgtcaactcacttgcagtgatggtgtaaacacacttcttcaaacacacttgaacac 1613
 QY 1681 tttttgtcaactcacttgcagtgatggtgtaaacacacttcttcaaacacacttgaacac 1740
 Db 1614 tttttgtcaactcacttgcagtgatggtgtaaacacacttcttcaaacacacttgaacac 1673
 QY 1741 ctcttgacccctcctcctcaatgcttctgagcctcactcctcctccttcttctgggctgt 1800
 Db 1674 ctcttgacccctcctcctcaatgcttctgagcctcactcctcctccttcttctgggctgt 1733

QY 1801 attacagccctctgtgatacttcaactctgtgctcctcactgtgagcgagctcaact 1860
 Db 1734 attacagccctctgtgatacttcaactctgtgctcctcactgtgagcgagctcaact 1793
 QY 1861 gtaactgacagtggtcgtcctctcttggcagatgatacaacactgtgaagtaacttaact 1920
 Db 1794 gtaactgacagtggtcgtcctctcttggcagatgatacaacactgtgaagtaacttaact 1853
 QY 1921 gcccaagcctgggagatcaggaaggtctgcataagctagtaagttgagttgacttgt 1980
 Db 1854 gcccaagcctgggagatcaggaaggtctgcataagctagtaagttgagttgacttgt 1913
 QY 1981 gtgtgatacagtgacttaagatctgttaataacttaattgtaatacagacactgttt 2040
 Db 1914 gtgtgatacagtgacttaagatctgttaataacttaattgtaatacagacactgttt 1973
 QY 2041 ttaaaccaagtaagactgtgtaaac 2069
 Db 1974 ttaaaccaagtaagactgtgtaaac 2002

RESULT 3
 AAF18148
 ID AAF18148 standard; DNA; 1645 BP.
 AC AAF18148;
 XX
 DT 14-MAR-2001 (first entry)
 XX
 DE Lung cancer associated polynucleotide sequence SEQ ID 167.
 XX
 KW Human; lung cancer associated protein; neuroprotective; cytostatic;
 KW cardioactive; immunomodulatory; muscular active; vulnerary;
 KW gastrointestinal; nephrotoxic; antilinfecive; gynecological;
 KW antibacterial; diagnosis; neural disorder; immune disorder; reproductive;
 KW proliferative disorder; wound healing; infectious disease; ds.
 XX
 OS Homo sapiens.
 XX
 PN WO20005180-A2.
 XX
 PD 21-SEP-2000.
 XX
 PF 08-MAR-2000; 2000MO-US05918.
 XX
 PR 12-MAR-1999; 99US-0124270.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 PA (ROSE/) ROSEN C A.
 PA
 PI Ruben SM;
 XX
 DR WPI; 2000-587514/55.
 DR P-PSDB; AAB58272.
 XX
 PT Lung cancer associated gene sequences, referred to as lung cancer
 PT antigens, useful for treatment, prevention, and diagnosis of disorders
 PT such as lung cancer -
 XX
 PS Claim 1; Page 631; 1425pp; English.
 XX
 CC Polynucleotide sequences AAF17982 - AAF18424 encode human lung cancer
 CC associated proteins represented in AAB58106 - AAB58548. Lung cancer
 CC associated proteins and polynucleotide sequences, their agonists, and
 CC antagonists may have neuroprotective; cytostatic; cardioactive;
 CC immunomodulatory; muscular active general; vulnerary; gastrointestinal
 CC general; nephrotoxic; antilinfecive; gynecological; or antibacterial
 CC activity. The invention also includes antibodies specific for the
 CC protein or polynucleotide sequences. The lung cancer associated
 CC polynucleotide sequences may be used for detection of lung cancer,
 CC chromosome identification, as chromosome markers, and for numerous other
 CC diagnostic or research purposes. The proteins may be used to treat

CC disorders such as neural, immune, muscular, reproductive,
CC gastrointestinal, pulmonary, cardiovascular, renal, and proliferative
CC disorders. The proteins may also be used in the treatment of wounds and
CC infectious diseases. Polynucleotide sequences AAF18425 - AAF18433 and
CC peptide AAB58549 are used in the course of the invention for the
CC identification and characterisation of the polynucleotide and protein
CC sequences.

XX
XX
SQ Sequence 1645 BP: 377 A; 386 C; 408 G; 468 T; 6 other:

Query Match 62.0%; Score 1312.8; DB 21; Length 1645;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1325; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

QY 770 gaaagccttgaaacggcgacatgtctctcaacatctacacgtctctctctctgtttgagaca 829
DB 297 gaaagccttgaaacggcgacatgtctctcaacatctacacgtctctctctctgtttgagaca 356
QY 830 tgcgttgaaagactatatacaactgtatcacaactgtgacaagcgtttgacaagctgag 889
DB 357 tgcgttgaaagactatatacaactgtatcacaactgtgacaagcgtttgacaagcgtgag 416
QY 890 gattggccagctgactgtctcagaagtaaaatcggctgttatactccgttgcctgaattc 949
DB 417 gattggccagctgactgtctcagaagtaaaatcggctgttatactccgttgcctgaattc 476
QY 950 taccgatattgttgaagataagtgctcgaactgtgctgggagaatgtgcgacaattctctg 1009
DB 477 taccgatattgttgaagataagtgctcgaactgtgctgggagaatgtgcgacaattctctg 536
QY 1010 cagccctgaaagagactggaactgtctctctctgctgcacaacgttgcctcctcctccaca 1069
DB 537 cag-cctgaaagagactggaactgtctctctctctgctgcacaacgttgcctcctcctccaca 595
QY 1070 ctacacttgagacaagcgcttgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 1129
DB 596 ctacacttgagacaagcgcttgaaatgaaatgaaatgaaatgaaatgaaatgaaatgaaat 655
QY 1130 tagtataagcagtgcaagaaacaaacagcacttgcagcaagctctgtgttactgttaact 1189
DB 656 tagtataagcagtgcaagaaacaaacagcacttgcagcaagctctgtgttactgttaact 715
QY 1190 gttgtgagagcagaag 1249
DB 716 gttgtgagagcagaag 775
QY 1250 aactgcagaacatctctcacaagctagaactgaacaaacccctctgtctagggtgtgtcc 1309
DB 776 aactgcagaacatctctcacaagctagaactgaacaaacccctctgtctagggtgtgtcc 835
QY 1310 gttgtgagagcagatctctctctcctcctataataataataataataataataataata 1369
DB 836 gttgtgagagcagatctctctcctcctcctataataataataataataataataataata 895
QY 1370 tactgggacttactgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 1429
DB 896 tactgggacttactgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtatgtat 955
QY 1430 gttgggaatcctatctgtgcttagactgtatccactcactcactcactcactcactcact 1489
DB 956 gttgggaatcctatctgtgcttagactgtatccactcactcactcactcactcactcact 1015
QY 1490 ttgt 1549
DB 1016 ttgt 1075
QY 1550 ctgtgtcactgactgactgt 1609
DB 1076 ctgtgtcactgactgactgt 1135
QY 1610 aagtggtttgttttttgttcaacttgcactgagcagtgagcagcagcagcagcagcagcagc 1669
DB 1610 aagtggtttgttttttgttcaacttgcactgagcagtgagcagcagcagcagcagcagcagc

DB 1136 aagtggtttgttttttgttcaacttgcactgagcagtgagcagcagcagcagcagcagcagc 1195
QY 1670 tctbaacaccccttctgactgagtgacttgaagaaatcttactatggaagaaatgaacag 1729
DB 1196 tctbaacaccccttctgactgagtgacttgaagaaatcttactatggaagaaatgaacag 1255
QY 1730 aagtgagtgcttcttgaaccccttccactcaatgttctgaacttcaacttccacttctct 1789
DB 1256 aagtgagtgcttcttgaaccccttccactcaatgttctgaacttcaacttccacttctct 1315
QY 1790 ttctggctgtatataagcagcctctgtgacttcaacttgcctgcctcactgtgagca 1849
DB 1316 ttctggctgtatataagcagcctctgtgacttcaacttgcctgcctcactgtgagca 1375
QY 1850 gcaatccacgtgactgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1909
DB 1376 gcaatccacgtgactgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgagtgag 1435
QY 1910 tactaatctgcccagcgttgggagagatcaggaaggtctgtcatagttagttagttagttagtt 1969
DB 1436 tactaatctgcccagcgttgggagagatcaggaaggtctgtcatagttagttagttagttagtt 1495
QY 1970 ttactgttgt 2029
DB 1496 ttactgttgt 1555
QY 2030 aagcagcttctttaaaccacagtaagactgtgaaacgtgtgtgtgtgtgtgtgtgtgtgtgt 2089
DB 1556 aagcagcttctttaaaccacagtaagactgtgaaacgtgtgtgtgtgtgtgtgtgtgtgtgt 1615
QY 2090 aaaaaaaanaa 2102
DB 1616 aaaaaaaanaa 1628

RESULT 4

AAX30152
ID AAX30152 standard; DNA; 1197 BP.

XX
XX
AC AAX30152;
XX
DT 18-JUN-1999 (first entry)
XX
DE Human secreted protein gene 8.
XX
KW Human; secreted protein; cancer; tumour; developmental abnormality;
KW foetal deficiency; blood disorder; immune system disorder; inflammation;
KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
KW digestive disorder; endocrine disorder; infection; AIDS; ss.
XX
OS Homo sapiens.
XX
PN W09910363-A1.
XX
PD 04-MAR-1999.
XX
XX
PF 27-AUG-1998; 98WO-US17709.
XX
PR 29-AUG-1997; 97US-0056271.
XX
PR 29-AUG-1997; 97US-0056073.
XX
PR 29-AUG-1997; 97US-0056247.
XX
PR 29-AUG-1997; 97US-0056270.
XX
PA (HDMA-) HUMAN GENOME SCI INC.
XX
PI Fan P, Kyaw H, Rosen CA, Ruben SM, Wei YF;
XX
DR WPI: 1999-190585/16.
XX
DR P-PSDB: AAV04300.
XX
PT New isolated human genes and the secreted polypeptides they encode

PT - useful for diagnosis and treatment of e.g. cancers, neurological disorders, immune diseases, inflammation or blood disorders

PS Claim 1; Page 136; 170pp; English.

AAV030145 to AAV030149 represent 29 isolated human secreted protein genes
CC AAV04293 to AAV04321 represent the secreted proteins encoded by the 29
CC human genes. The genes and their corresponding secreted polypeptides are
CC useful for preventing, treating or ameliorating medical conditions,
CC e.g. by protein or gene therapy. Also pathological conditions can be
CC diagnosed by determining the amount of the new polypeptides in a sample
CC or by determining the presence of mutations in the new genes. Specific
CC uses are described for each of the 29 genes, based on which tissues they
CC are most highly expressed in, and include developing products for the
CC diagnosis or treatment of cancer, tumours, developmental abnormalities
CC and foetal deficiencies, blood disorders, diseases of the immune system,
CC autoimmune diseases, inflammation, allergies, Alzheimer's and cognitive
CC disorders, schizophrenia, arthritis, asthma, psoriasis, sepsis, skin
CC disorders, atherosclerosis, diabetes, cardiovascular disorders, kidney
CC disorders, digestive/endocrine disorders, infections and AIDS. The
CC polypeptides are also useful for identifying their binding partners.
CC The sequences given in AAV030174 to AAV030182 and AAV04334 are
CC used in the exemplification of the present invention.

SQ Sequence 1197 BP; 289 A; 270 C; 287 G; 345 T; 6 other;

Query Match	51.78; Score 1095.6; DB 20; Length 1197
-------------	---

Matches 1101; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

[illegible]

Db	666	tcctcttcgataagccaaataacgaagtgttggtttttttgtacccttcactgcatg	725
Qy	1646	gctaacacacttttttttccaacccctgtacaacactttttctgcatggtgaacttgcaggaa	1705
Db	726	gttaaccacactttttttccaacccctgtacaacccctttttctgattggtgaacttgcaggaa	785
Qy	1706	tattctatttgaaaagataacagagatcaagtgctctcttgacccttcctcaatgttt	1765
Db	786	tattctatttgaaaagataacagagatcaagtgctctcttgacccttcctcaatgttt	845
Qy	1766	ctagccttcacttcacatgtctctttctctggtcgtatatacagccctctgcatctcaa	1825
Db	846	ctagccttcacttcacatgtctctttctctggtcgtatatacagccctctgcatctcaa	905
Qy	1826	ctctctgcccctccacttgatgcatgacagatccaacagttaactgacagtggtcgctctct	1885
Db	906	ctctctgcccctccacttgatgcatgacagatccaacagttaactgacagtggtcgctctct	965
Qy	1886	ggggccatgcatcacacccctgtaaagtactcaatctactgccacgcttggggagatacagagag	1945
Db	966	ggggccatgcatcacacccctgtaaagtactcaatctactgccacgcttggggagatacagagag	1025
Qy	1946	gtctgcatagttgaagaattgggttaactctttgtgtgtgcatcagtgacttagagttct	2005
Db	1026	gtctgcatagttgaagaattgggttaactctttgtgtgtgcatcagtgacttagagttct	1085
Qy	2006	gtaataacttattgtaaatgcatgacagacgtttttaaccocaaagataagactgtctga	2065
Db	1086	gtaataacttattgtaaatgcatgacagacgtttttaaccocaaagataagactgtctga	1145
Qy	2066	aacctgttgatgtaaaaaaataaaaaaataaaaaa	2100
Db	1146	aacctgttgatgtaaaaaaataaaaaaataaaaaa	1180

RESULT	5
AAH03869	
ID	AAH03869 standard; cDNA: 783 BP.
XX	AAH03869;
AC	
XX	
DT	26-JUN-2001 (first entry)
XX	
DE	Human cDNA clone (5'-primer) SEQ ID NO:704.
XX	
KW	Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.
KM	
XX	
OS	Homo sapiens.
XX	
PN	EP1074617-A2.
XX	
PD	07-FEB-2001.
XX	
PF	28-JUL-2000; 2000EP-0116126.
XX	
PR	29-JUL-1999; 99JP-0248036.
PR	27-AUG-1999; 99JP-0300253.
PR	11-JAN-2000; 2000JP-0118776.
PR	02-MAY-2000; 2000JP-0183767.
XX	
XX	09-JUN-2000; 2000JP-0241899.
PA	(HELI-) HELIX RES INST.
XX	
PI	Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI	Isihl S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX	
DR	WPI: 2001-318749/34.
XX	
PT	Primer sets for synthesizing polynucleotides, particularly the 5602
PT	full-length cDNAs defined in the specification, and for the detection
PT	and/or diagnosis of the abnormality of the proteins encoded by the
PT	full-length cDNAs -
XX	

OY 181 ctatgtgttaaccagacaaagcattggaataaccatcgttcttgggtcgc 240
 |||||
 DB 204 ctatgtgttaaccagacaaagcattggaataaccatcgttcttgggtcgc 263
 OY 241 attccccgctctatcctgttacacaatcttcataagaa 284
 |||||
 DB 264 attccccgctctatcctgttacacaatcttcataagaa 307

RESULT 10

AAF93295
 ID AAF93295 standard; cDNA; 274 BP.

AC AAF93295;
 XX
 DT 21-MAY-2001 (first entry)

DE Lung carcinoma cDNA encoding SRT protein SEQ ID 116.
 XX
 KM Human; SRT; gene therapy; gene mapping; tissue typing; ss.

XX Homo sapiens.
 OS
 PN WO200107611-A2.

PD 01-FEB-2001.
 XX
 PF 21-JUL-2000; 2000MO-US20006.

XX 26-JUL-1999; 990S-0145701.
 XX
 PA (GETH) GENENTECH INC.

PI Baker KP, Goddard A, Wood WI;
 XX
 DR WPI: 2001-112729/12.

XX New isolated nucleic acid molecule encoding a SRT polypeptide is useful
 PT for production of recombinant SRT polypeptides, gene mapping,
 PT diagnosing genetic disorders and for gene therapy -
 XX
 PS Claim 2; Fig 116; 663pp; English.

XX Sequences AAF93180 - AAF93743 represent polynucleotide sequences encoding
 CC human SRT proteins. The cDNA sequences are isolated from various
 CC different human tissue cDNA libraries. The invention relates to a method
 CC for detecting cDNA encoding an SRT protein, a vector containing cDNA
 CC encoding SRT, a host cell transformed with the vector, an isolated SRT
 CC polypeptide, and an antibody which binds to SRT. The polynucleotide
 CC sequence can be used in gene therapy and is useful in the recombinant
 CC production of SRT polypeptides, as a hybridisation probe to screen
 CC libraries to isolate cDNAs with sequence identity to SRT polypeptides, to
 CC map the gene encoding the SRT polypeptides and analysing genetic
 CC disorders, tissue typing and disease detection. The SRT
 CC polynucleotide sequences can be used in polymerase chain reaction,
 CC screening for new therapeutic molecules and generation of antisense RNA
 CC and DNA.
 XX
 PS Sequence 274 BP; 79 A; 53 C; 52 G; 84 T; 6 other;

Query Match 12.2%; Score 258.4; DB 22; Length 274;
 Best Local Similarity 95.6%; Pred. No. 2,4e-54;

Matches 262; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 238 gtcatctcccgctctatcctgttaacacaatcttcagaagaaggttcagatgtat 297
 |||||
 DB 1 gtcatctcccgctctatcctgttaacacaatcttcagaagaaggttcagatgtat 60

OY 298 gggcgtatgcaaaaagcctagaagaataagacaataatgtggaagcacataaagt 357
 |||||
 DB 61 gggcgtatgcaaaaagcctagaagaataagacaataatgtggaagcacataaagt 120

OY 358 ttcatcaatccatccagcagagatgagacatttgagacaggttccgcgaagctcacca 417
 |||||
 DB 121 ttatcaantccatccagcagagatgagacatttgagacaggttccgcgaagctcacca 180
 OY 418 agtgccttctccatgattatcttcacatccacttttgcaactactcgtctctgc 477
 |||||
 DB 181 agtgccttctccatgattatcttcacatccacttttgcaactactcgtctctgc 240
 OY 478 taatgacctgttcccgagcaactatgatcag 511
 |||||
 DB 241 taatgacctgttcccgagcaactatgatcag 274

RESULT 11

AAH30637
 ID AAH30637 standard; cDNA; 398 BP.

AC AAH30637;
 XX
 DT 27-JUL-2001 (first entry)

DE Human colon cancer cell line Km12L4-A cDNA library derived sequence #571.
 XX
 KM Human; diagnosis; colon cancer; cancer; malignant; chromosome mapping;
 KM detection; colon cancer cell line Km12L4-A; ss.
 XX
 OS Homo sapiens.

PN WO200018916-A2.
 XX
 PD 06-APR-2000.

PF 23-SEP-1999; 99WO-US222226.
 XX
 PR 28-SEP-1998; 98US-0102161.

XX 28-SEP-1998; 98US-0102180.
 PR 29-SEP-1998; 98US-0102380.
 PR 08-OCT-1998; 98US-0103815.
 PR 27-OCT-1998; 98US-0105877.

PA (CHIR) CHIRON CORP.
 PA (HYSE-) HYSEQ INC.

PI Williams LT, Escobedo J, Innis MA, Garcia PD, Sudduth-Klinger J;
 PI Reinhard C, Giese K, Randazzo F, Kennedy GC, Pot D, Kassam A;
 PI Lamson G, Drmanac R, Crkvenjakov R, Dickson M, Drmanac S, Labat I;
 PI Leshkowitz D, Kita D, Garcia V, Jones LW, Stache-Crain B;
 DR WPI: 2000-293155/25.

XX Polynucleotide library comprising 1079 defined sequences, useful in
 PT the form of an array to detect cancer or susceptibility to cancer -
 XX
 PS Claim 1; Page 347; 502pp; English.

XX The present invention describes a library of polynucleotides comprising
 CC 1079 nucleotide sequences (given in AAH30637 to AAH31145). Also described
 CC are: (1) an isolated polynucleotide (I) having at least 90% identity to
 CC one of the 1079 sequences; (2) a recombinant host cell containing (1);
 CC (3) an isolated polypeptide (II) encoded by (1); (4) an antibody that
 CC specifically binds to (II); (5) a vector comprising (I); and (6) a method
 CC of detecting differentially expressed genes correlated with a cancerous
 CC state of a mammalian cell comprising detecting a gene product encoded by
 CC 65 of the 1079 sequences given in the specification. The polynucleotides
 CC are used to monitor patients having (or susceptible) to cancer to detect
 CC potentially malignant events at a molecular level before they are
 CC detectable at a gross morphological level. The polynucleotides are also
 CC useful for monitoring the efficacy of various therapies and preventive
 CC interventions. Polynucleotide probes based on the disclosed sequences
 CC are useful for chromosome mapping and detection of transcription levels.
 CC The 1079 polynucleotide sequences were derived from a human colon cancer
 CC cell line Km12L4-A cDNA library.
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:10:52 ; Search time 13.29 Seconds

(without alignments)
993.178 Million cell updates/sec

Title: US-09-868-474-2

Sequence: 1 MALSRCVMSASVWGSATVP.....LSLLHNVLSTNYLSTR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171.5	9.0	1013	1 A60D_DROME	P91927 drosophila
2	92	4.8	656	1 SYM_HELPJ	Q92K99 helicobacte
3	91.5	4.8	1120	1 YB85_YEAST	P38329 saccharomyc
4	89	4.7	276	1 D101_ARATH	Q05211 arabidopsis
5	86	4.5	3066	1 POLG_BCMVN	Q65399 b genome po
6	85.5	4.5	2231	1 SENL_YEAST	Q00416 saccharomyc
7	84	4.4	561	1 SYR_CHLPP	Q927Y3 chlamydia p
8	84	4.4	650	1 SYM_HELPJ	P56127 helicobacte
9	83.5	4.4	603	1 UVRG_CHLPP	Q926W6 chlamydia p
10	83.5	4.4	738	1 NU5C_LIGVU	P87143 ligustrum v
11	82.5	4.3	2199	1 DPOE_SCHPO	Q91423 schizosach
12	81.5	4.3	816	1 AD15_RAT	Q9G9V0 r adam 15 p
13	81.5	4.3	1230	1 PRPM_HUMAN	Q00411 homo sapien
14	81	4.2	1373	1 HRP1_SCHPO	Q94525 schizosach
15	81	4.2	1447	1 TOP2_DROME	P15348 drosophila
16	79.5	4.2	573	1 CYDC_ECOLI	P23886 escherichia
17	79.5	4.2	2215	1 MY7A_HUMAN	Q13402 homo sapien
18	79	4.1	500	1 Y039_BORBU	Q51068 borrelia bu
19	79	4.1	1181	1 HAIR_RAT	P97609 rattus notv
20	79	4.1	1182	1 HAIR_MOUSE	Q61645 mus musculu
21	78.5	4.1	469	1 P49_STR1	P06108 streptomyce
22	78.5	4.1	505	1 PR12_MOUSE	P36310 mus musculu
23	78	4.1	536	1 CCA4_DROME	Q9VE00 drosophila
24	78	4.1	800	1 PT11_YEAST	P08468 saccharomyc
25	78	4.1	847	1 HEX_VIBVU	Q04786 vibrio vuln
26	77.5	4.1	310	1 HTPX_HELPJ	Q25582 helicobacte
27	77.5	4.1	383	1 HOXV_ALCEU	P31914 alcaligenes
28	77.5	4.1	398	1 PARA_ECOLI	P07620 escherichia
29	77.5	4.1	505	1 RRG_NOTVI	P18516 notophthalam
30	77.5	4.1	2747	1 FAF_DROME	P55824 drosophila
31	77	4.0	312	1 Y003_BPL2	P42538 bacteriophia
32	77	4.0	708	1 YB40_HUMAN	Q94100 homo sapien
33	77	4.0	873	1 COX1_HUMAN	Q37370 acanthamoeb

34	76.5	4.0	206	1 ICT1_HUMAN	Q14197 homo sapien
35	76.5	4.0	346	1 NU2M_CORJA	P24971 coturnix co
36	76.5	4.0	436	1 UI32_HSVSA	Q01040 herpesvirus
37	76.5	4.0	500	1 PTG1_HUMAN	Q16647 homo sapien
38	76.5	4.0	722	1 GRE4_CHLPP	Q92794 chlamydia p
39	76.5	4.0	842	1 ABC6_HUMAN	Q9NP58 homo sapien
40	76.5	4.0	1878	1 BRCL_CANPA	Q95153 canis famli
41	76.5	4.0	2472	1 SPCN_HUMAN	Q13813 homo sapien
42	76	4.0	380	1 KPT3_HUMAN	Q07002 homo sapien
43	76	4.0	1067	1 ST1_PYRAB	Q9V072 pyrococcus
44	75.5	3.9	310	1 HTPX_HELPJ	Q92K54 helicobacte
45	75.5	3.9	452	1 GP13_YEAST	P32363 saccharomyc

ALIGNMENTS

RESULT	ID	Accession	Standard	PRT	AA
1	A60D_DROME	Q91927	Q91927	1013	AA
AC	P91927	Q91927	Q91927	1013	AA
DT	30-MAY-2000	(Rel. 39, Created)			
DT	20-AUG-2001	(Rel. 40, Last sequence update)			
DE	20-AUG-2001	(Rel. 40, Last annotation update)			
DE	CALCIUM-BINDING MITOCHONDRIAL PROTEIN ANON-60DA.				
GN	ANON-60DA OR CG4589				
OS	Drosophila melanogaster (Fruit fly).				
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;				
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;				
OC	Ephydroidea; Drosophilidae; Drosophila.				
OX	NCBI_TaxID=7227;				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=BERKELEY;				
RC	MEDLINE=20196006; PubMed=10731132;				
RA	Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,				
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,				
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,				
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,				
RA	Brandon R.C., Rogers J.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,				
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,				
RA	Abrial J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,				
RA	Ballev R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,				
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,				
RA	Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,				
RA	Buttis K.C., Busam D.A., Butler H., Cadenot L.B., Davies A., Chandra I.,				
RA	Cherry J.W., Cawley S., Dahlke C., Davenport L.B., Davies P.,				
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,				
RA	Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,				
RA	Doulin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,				
RA	Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,				
RA	Glocke A., Gong F., Gottrell J.H., Gu Z., Guan P., Harris M.,				
RA	Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,				
RA	Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,				
RA	Jaitani B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,				
RA	Kimmel M.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,				
RA	Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang X., Lin X.,				
RA	Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,				
RA	Metkulov G., Milshina N.V., Mobarry C., Morris J., Moshier A.,				
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,				
RA	Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,				
RA	Palazzo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,				
RA	Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,				
RA	Shue B.C., Sider-Klimos I., Simpson M., Skupski M.P., Smith T.,				
RA	Spieler E., Spradling A.C., Stapleton M., Strong R., Sun E.,				
RA	Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,				
RA	Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,				
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,				
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,				
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,				
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;				
RT	"The genome sequence of Drosophila melanogaster".				
RL	Science 287:2185-2195(2000).				

```

RN [2]
RP SEQUENCE OF 626-944 FROM N.A.
RC TISSUE-Ovary:
RX MEDLINE-99168769: Pubmed-10071211:
RA Caggese C., Ragone G., Perrini B., Moschetti R., de Pinto V.,
RA Calzati R., Barsanti P.:
RT "Identification of nuclear genes encoding mitochondrial proteins:
RT isolation of a collection of D. melanogaster cDNAs homologous to
RT sequences in the Human Gene Index database."
RL Mol. Gen. Genet. 261:64-70(1999).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO
CC FRAMESHIFTS IN POSITIONS 920 AND 930.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
DR EMBL: AE003464; AA047217.1; -
DR EMBL: Y10912; CAAT1853.1; ALT_FRAME.
DR Flybase: FBgn0019886; CG4598.
DR InterPro: IPR001813; 60s.ribosomal.
DR InterPro: IPR002024; Bacterioferritin.
DR InterPro: IPR003871; DUF223.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000861; REM_repeat.
DR InterPro: IPR002555; RFL.
DR InterPro: IPR001236; 1dh.
DR Pfam: PF00428; 60s.ribosomal; 1.
DR Pfam: PF01334; Bacteriofer; 1.
DR Pfam: PF02721; DUF223; 1.
DR Pfam: PF00036; ehand; 2.
DR Pfam: PF02185; HRI; 1.
DR Pfam: PF00056; 1dh; 1.
DR Pfam: PF01605; RFL; 1.
DR SMART: SM00054; Eph; 2.
DR PROSITE: PS00018; EF_HAND; 2.
DR Mitochondrion: Calcium-binding; Repeat.
DR CA_BIND 700 711 EF_HAND 1 (POTENTIAL).
DR CA_BIND 783 794 EF_HAND 2 (POTENTIAL).
DR CONFLICT 687 687 E -> Q (IN REF. 2).
DR CONFLICT 690 690 K -> G (IN REF. 2).
DR CONFLICT 717 717 Q -> H (IN REF. 2).
DR CONFLICT 736 736 E -> D (IN REF. 2).
DR CONFLICT 740 740 K -> R (IN REF. 2).
DR CONFLICT 761 761 K -> E (IN REF. 2).
DR CONFLICT 943 943 G -> S (IN REF. 2).
SQ SEQUENCE 1013 AA; 113579 MW; 8C710380263F262E CRC64;

```

```

Query Match 9.0%; Score 171.5; DB 1; Length 1013;
Best Local Similarity 21.7%; Pred. No. 2.1e-07;
Matches 66; Conservative 54; Mismatches 119; Indels 65; Gaps 8;

```

```

QY 100 AKKARRITNMKKNIKEHQ-----LPRREHMLRQFQDV 135
      || : : : : : : : : : : : : : : : : : : : :
DB 175 AKPKPRTRTMDLVHYHGFRLLEFDVAICSLKRLRWLNGKTLTRRKNOLORTTSDL 234
      || : : : : : : : : : : : : : : : : : : : :
QY 136 TKCFELGISIPRPANLVFLMYLFP-----ROLLRHRWTK-----QQ 176
      || : : : : : : : : : : : : : : : : : : : :
DB 235 FRLLPFSVFTIVPEMLLPLFTKFPQEMLPSTQTSTDRQKIKOSLSVRLVAKFLQQ 294
      || : : : : : : : : : : : : : : : : : : : :
QY 177 T-DELDIYHAFKOSHEPILSYLEKVIPLISDAGLRWLTDCTKIORGTHPAIHD-ILA 234
      || : : : : : : : : : : : : : : : : : : : :
DB 295 TLDQMPVGH---KHSSEAKQFEAF-----TKINPPEPVSNDELITK 335
      || : : : : : : : : : : : : : : : : : : : :
QY 235 LRECFSNHPLGMNOLQALHVAKLSRAMLTSYLPPLLRHRLKTHVTTHQDLAKLKG 294
      || : : : : : : : : : : : : : : : : : : : :

```

```

DB 336 FAKRPDDE-ITLDSLSREOLAACRYLEINTIGTTTLRFQRLKRSIATDDRIAREG 394
      || : : : : : : : : : : : : : : : : : : : :
QY 295 IGLTAQEVKSCACYLGRGLNSTHIGEDRCRTWLGEMLOISCSLKEAELSLHNVLLSTN 354
      || : : : : : : : : : : : : : : : : : : : :
DB 395 VBSLDLELQQAQKARGMAYGITEERLRFLKEMIDLSINQVPTPTLLLSRTMLISDD 454
      || : : : : : : : : : : : : : : : : : : : :
QY 355 YLGT 358
      || : : : : : : : : : : : : : : : : : : : :
DB 455 SIT 458

```

RESULT 2

```

SYM_HELPJ
ID SYM_HELPJ STANDARD; PRT; 656 AA.

```

```

AC 092KG9;

```

```

DT 30-MAY-2000 (rel. 39, Created)

```

```

DT 30-MAY-2000 (rel. 39, Last sequence update)

```

```

DT 20-AUG-2001 (rel. 40, Last annotation update)

```

```

DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)

```

```

DE (METS).

```

```

GN METG OR JHP0967.

```

```

OC Helicobacter pylori J99 (Campylobacter pylori J99).

```

```

OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;

```

```

OX NCBI_TaxID=85963;

```

```

RP SEQUENCE FROM N.A.

```

```

RX MEDLINE-99120557: Pubmed-9923682;

```

```

RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Dolg P.C.,

```

```

RA Smith D.R., Noonan B., Gull B.C., deGange B.L., Carmel G.,

```

```

RA Tumundo P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,

```

```

RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Voyis G.F.,

```

```

RA Trust T.J.;

```

```

RT "Genomic sequence comparison of two unrelated isolates of the human

```

```

RT gastric pathogen Helicobacter pylori."

```

```

RL Nature 397:176-180(1999).

```

```

CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING

```

```

CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO

```

```

CC FOR THE INITIATION OF ALL RNA TRANSLATION THROUGH INITIATOR

```

```

CC TRNA(EMET) AMINOACYLATION.

```

```

CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +

```

```

CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).

```

```

CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).

```

```

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.

```

```

CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.

```

```

CC STRONG, TO CYSTEINYL-TRNA SYNTHETASE.

```

```

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----

```

```

CC EMBL: AE001525; AAD06543.1; -

```

```

CC InterPro: IPR002300; trna-synt_1a.

```

```

CC InterPro: IPR001412; trna-synt_1.

```

```

CC InterPro: IPR002304; trna-synt_met.

```

```

CC InterPro: IPR002547; trna_bind.

```

```

CC Pfam: PF00133; trna-synt_1; 1.

```

```

CC Pfam: PF01588; trna_bind; 1.

```

```

CC PRINTS: PR01041; TRNASYNTMET.

```

```

DR PROSITE: PS00178; AA TRNA LIGASE I; FALSE NEG.

```

```

KW Aminoacyl-TRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

```

```

KW trna-binding; Complete proteome.

```

```

KW SITE 11 21 "HIGH" REGION.

```

```

KW SITE 301 305 "KMSK" REGION.

```

```

KW BINDING 304 304 ATP (BY SIMILARITY).

```

```

KW DOMAIN 561 654 TRNA BINDING.

```

```

SQ SEQUENCE 656 AA; 75433 MW; 8F22F3BA1F9B9756 CRC64;

```

FT	TRANSMEM	321	341	POTENTIAL.
FT	TRANSMEM <td>361</td> <td>381</td> <td>POTENTIAL.</td>	361	381	POTENTIAL.
FT	TRANSMEM <td>394</td> <td>414</td> <td>POTENTIAL.</td>	394	414	POTENTIAL.
FT	TRANSMEM <td>431</td> <td>451</td> <td>POTENTIAL.</td>	431	451	POTENTIAL.
FT	TRANSMEM <td>482</td> <td>502</td> <td>POTENTIAL.</td>	482	502	POTENTIAL.
FT	TRANSMEM <td>598</td> <td>618</td> <td>POTENTIAL.</td>	598	618	POTENTIAL.
FT	TRANSMEM <td>620</td> <td>640</td> <td>POTENTIAL.</td>	620	640	POTENTIAL.
FT	TRANSMEM <td>755</td> <td>775</td> <td>POTENTIAL.</td>	755	775	POTENTIAL.
SO	SEQUENCE	1120 AA:	123998 MW:	299699C0CB3E5CC CRC64;

Query Match	4.8%;	Score 91.5;	DB 1;	Length 1120;
Best Local Similarity	19.5%;	Pred. No. 3.2;		
Matches 68;	Conservative 48;	Mismatches 120;	Indels 113;	Gaps 16

QY	78	HFPEFYLYTIFPKKGOMLADAKKARIKTYMMKHNKIFHQLPYREMEHLROFROOVTK	137
Db	403	HSPLYSIFFWLITQL-CLESDVNKIAITIT-----MTFLMTFVNNLA	445
QY	138	CLFLGIIISIPF-----	165
Db	446	CFLLGISAPENFSPSEKRYENRYTATIGALLSVAMLIVDGISASYFLFAMILLFLF----	501
QY	166	LIRHFWTPKQOTFLD-LYHAFRK-----QSHPEIISYLE-KVIFLISDAGLRMTDLC	219
Db	502	-HYFSPKSGVSGVSSLLIYHOVKYLLRLRODNKIKYRPDILLFVDNPRTSMILRFEN	560
QY	220	KIORG-----THPAIHDLALRECFSSNHPLEGNLOQA-----LHVKALSRAMLTSTYL	267
Db	561	HLKKGGLYILGHVAV-----TADPEPKOLNEIKTQOKAMKIRDMAIKAFVQGTG	611
QY	268	PPPLLRH-----LKTHTYVYIHOIDKALALAGIGLTAQ-----EYK-----S	305
Db	612	PSLIMGIRNFIYSSGLGGMKPNITVYGFEDLESTYKHLIPDSRSNNIIOKOVETIKATVP	671
QY	306	ACYLRGLNSTHGEDRCRTW---LGEMLQISCSLKEAELSLLHNVYL	350
Db	672	TCSDVAKIN-VLPETDECKNETKVNQOWOI-----VEDLSLMQSNIAI	714

RESULT 4			
ID	D101_ARATH	STANDARD:	PRT: 276 AA.
AC	005211;		
DT	01-NOV-1995 (Rel. 32, Created)		
DT	01-NOV-1995 (Rel. 32, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	DNA-DAMAGE-REPAIR/TOLERATION PROTEIN DRT101 PRECURSOR.		
GN	DRT101.		
OS	Arabidopsis thaliana (Mouse-ear cress).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;		
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsids.		
OX	NCBI_TaxID=3702;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=CV. COLUMBIA;		
RA	MEDLINE=93320380; PubMed=8329681;		
RA	Pang O.-, Hays J.B., Rajagopal I., Schaefer T.S.;		
RT	"Selection of Arabidopsis cDNAs that partially correct phenotypes of		
RT	Escherichia coli DNA-damage-sensitive mutants and analysis of two		
RT	plant cDNAs that appear to express UV-specific dark repair		
RT	activities.";		
RL	Plant Mol. Biol. 22:411-426(1993).		
CC	-1-SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST		
CC	PHOSPHOACETYLGLUCOSAMINE MUTASE (PCMI).		
CC			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
CC	the European Bioinformatics Institute. There are no restrictions on ways		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement		

or send an email to license@lsb.sib.ch.

EMBL: L11367; AAA72352.1; -

transit peptide; Chloroplast; DNA repair; DNA damage.

TRANSIT 1 ? CHLOROPLAST (POTENTIAL).

CHAIN ? 276 DNA-DAMAGE-REPAIR/TOLERATION PROTEIN

FT D8T101.

SEQUENCE 276 AA; 31063 MW; EC6FBD4061C06964 CRR64;

Query Match 4.7%; Score 89; DB 1; Length 276;

Best local Similarity 22.1%; Pred. No. 0.97;

Matches 78; Conservative 47; Mismatches 116; Indels 112; Gaps 21;

36 AKAPSSKILHLPRKADVKLMYVYVTKRKAIKNGHRLGRFPFYLITFMGLDM 95

5 SMMVRSIGFYSPSPDLKRL-SYLAV-----TFQCCLSSSSSNM 46

96 LMDAKARIKTNMKNHNIKFLQPLREMEHLRPFODVTKCLFLGISIP-PFANYLV 154

47 LMSMKKSSLLVMCR-----HLREM-----CVY-----RLPRAFETML 82

155 FLMLYLPROLLIRHFWTKQDTFLDIYHAFKQSHPELITYLEKVIPLISDAGLRML 214

83 FL-----KLEFKHLHEKAEFD-IGIY--FEANGHGT-----LFSESLMWLV 124

215 T-----DLCTKIQRTG--HRAHDLALRECSNPLRMNQLALHVKALSRAML----- 262

125 SKQDILFAKGGSGSEKHAVALAVSNLI-NOAVG-----DALSGVLLVEVILH 174

263 -----LTSYLPPLLRHLRKLTHTVYI--HOLDKALALGIGOLVAQEKSAQYLK 310

175 LGMSTKWNMLYKDLPSRQIKVEVPRTAVVTSSETEALRPNGIDAINSEIKK--YRS 232

311 GL-----NSTHIGEDRCRTWLGEMLOISGLKAELESLILHNVLLSTNYLGT 358

233 GRAFIRPSGT---EDVYRVY-----AAASTOE-DADSLANSVAOLVKSFIS 275

RESULT 5

ID POLG_BCMVN STANDARD; PRT; 3066 AA.

AC 065399;

DT 15-DEC-1998 (Rel. 37, Created)

DT 15-DEC-1998 (Rel. 37, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE GENOME POLYPROTEIN (CONTAINS: N-TERMINAL PROTEIN (P1); HELPER COMPONENT PROTEINASE (EC 3.4.22.-) (HC-PRO); PROTEIN P3; 6 KDA PROTEIN 1 (6K1); CYTOPLASMIC INCLUSION PROTEIN (CI); 6 KDA PROTEIN 2 (6K2); GENOME-LINKED PROTEIN (VPG); NUCLEAR INCLUSION PROTEIN A (NI-A) (NIA) (EC 3.4.22.44) (49 KDA PROTEINASE) (49 KDA-PRO); NUCLEAR INCLUSION PROTEIN B (NI-B) (NIB) (RNA-DIRECTED RNA POLYMERASE) (EC 2.7.7.48); COAT PROTEIN (CP)).

OS Bean common mosaic virus (strain NL-3 / Michigan) (BCMV).

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Potyviridae;

OC Potyvirus.

OX NCBI_TaxID=12196;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=96191623; PubMed=8607279;

RA Fang G.W., Allison R.F., Zambolim E.M., Maxwell D.P., Gilbertson R.L.;

RT "The complete nucleotide sequence and genome organization of bean common mosaic virus (NL3 strain).";

RL Virus Res. 39:13-23(1995).

CC -1- FUNCTION: HELPER COMPONENT-PROTEINASE IS REQUIRED FOR APHID TRANSMISSION AND ALSO HAS PROTEOLYTIC ACTIVITY.

CC -1- FUNCTION: CYTOPLASMIC INCLUSION PROTEIN HAS HELICASE ACTIVITY. IT MAY BE INVOLVED IN REPLICATION.

CC -1- FUNCTION: NUCLEAR INCLUSION PROTEIN A HAS PROTEOLYTIC ACTIVITY.

CC -1- CATALYTIC ACTIVITY: NUCLEAR INCLUSION PROTEIN A HYDROLASES GLUTAMINYL BONDS, AND ACTIVITY IS FURTHER RESTRICTED BY PREFERENCES FOR THE AMINO ACIDS IN P6 -P1 THAT VARY WITH THE SPECIES OF POTYVIRUS, E.G. GLU-XAA-XAA-TYR-XAA-GLN+ (SER OR GLY

FOR THE ENZYME FROM TOBACCO ETCH VIRUS. THE NATURAL SUBSTRATE IS THE VIRAL POLYPROTEIN, BUT OTHER PROTEINS AND OLIGOPEPTIDES CONTAINING THE APPROPRIATE CONSENSUS SEQUENCE ARE ALSO CLEAVED.

-1- PTH: VPG IS COVALENTLY LINKED TO THE GENOMIC RNA.

-1- PTH: THE VIRAL RNA OF POTYVIRUSES IS EXPRESSED AS A SINGLE POLYPROTEIN WHICH UNDERGOES POSTTRANSLATIONAL PROTEOLYTIC PROCESSING RESULTING IN THE PRODUCTION OF AT LEAST EIGHT INDIVIDUAL PROTEINS.

-1- SIMILARITY: HC PROTEINASE BELONGS TO PEPTIDASE FAMILY C6.

-1- SIMILARITY: NI-A PROTEINASE BELONGS TO PEPTIDASE FAMILY C4.

-1- SIMILARITY: BELONGS TO THE POTYVIRUSES POLYPROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.lsb.sib.ch/announce/>).

CC or send an email to license@lsb.sib.ch.

DR EMBL: U19287; AAB02170.1; -

DR MEROPS: C04.001; -

DR MEROPS: C06.001; -

DR MEROPS: S30.001; -

DR Interpro: IPR001410; DEAD.

DR Interpro: IPR001650; Helicase_C.

DR Interpro: IPR001730; Peptidase_C4.

DR Interpro: IPR001456; Peptidase_C6.

DR Interpro: IPR002540; Pcty_P1.

DR Interpro: IPR001592; Pcty_coat.

DR Interpro: IPR001205; RNA_pol_P3D.

DR Interpro: IPR001254; Trypsin.

DR Pfam: PF00863; Peptidase_C4; 1.

DR Pfam: PF00851; Peptidase_C6; 1.

DR Pfam: PF00767; Pcty_coat; 1.

DR Pfam: PF01577; Pcty_P1; 1.

DR Pfam: PF00680; RNA_dep_RNA_pol; 1.

DR PRINTS: PR00966; NIAPOTYPTASE.

DR SMART: SM00487; DEXDC; 1.

DR SMART: SM00490; HELICC; 1.

DR Hydrolase: Transferase; Thiol protease; RNA-directed RNA polymerase;

KW Coat protein; Polypeptide; Covalent protein-RNA linkage; Helicase;

KW ATP-binding.

KW CHAIN 1 ? N-TERMINAL PROTEIN.

FT CHAIN ? ? HELPER COMPONENT PROTEINASE.

FT CHAIN ? ? PROTEIN P3.

FT CHAIN ? ? 6 KDA PROTEIN 1.

FT CHAIN ? ? CYTOPLASMIC INCLUSION PROTEIN.

FT CHAIN ? ? 6 KDA PROTEIN 2.

FT CHAIN ? ? GENOME-LINKED PROTEIN.

FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN A.

FT CHAIN ? ? NUCLEAR INCLUSION PROTEIN B.

FT CHAIN ? ? COAT PROTEIN.

FT NP_BIND 1258 1265 ATP (POTENTIAL).

FT SEQUENCE 3066 AA; 350385 MW; E358955297FA359 CRR64;

Query Match 4.5%; Score 86; DB 1; Length 3066;

Best local Similarity 22.6%; Pred. No. 34;

Matches 56; Conservative 44; Mismatches 78; Indels 70; Gaps 15;

139 LFLGISIPFANYLVLLMYFPROLLIRH-----WTPKQGT---DFLDIYHAFKQ 189

808 LVGLVLS-----PTVLIHMF-----RMKHFKEGVLEMLNKDQSVVIFLLEHLTKRI 855

190 SHEPILSYLEKVIPLISDGLKRWRLDCTKIQRGTHP--AIHDLALR--ECESNHLG 245

856 ANMDVY--LEQLEMSQAG--RLHEILCDPKNHSTRAYKDELEVMEALTNKELA 910

246 MNDL-----QALVKALSRAMLTSYLPPLLRHLRKLTHTVYIHQ-LDKALAKL 293

911 NNGFPINSLGHVSKITAKALEKEMRALSWLESSVYWLKRSKYTEHLTKKAE- 969

QY 294 GIGGL-THOEVKASACYLGNLSTHGEDRCFTWLGEMLOISCKEAPLSLLHNVLIS 352
 Db 970 --GRKSSRRKFFVSACFMNA--QTHLGNAR-----ITISKNVEV 1004
 QY 353 TNYLGSTR 360
 Db 1005 TN-LGVR 1011

RESULT 6
 SENL_YEAST STANDARD: PRT: 2231 AA.
 AC 000416; 006448;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE TRNA-SPlicing ENDONUCLEASE POSITIVE EFFECTOR.
 GN SEN1 OR YIR430W OR L9576.1.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE OF 120-2231 FROM N.A.
 RX MEDLINE=92236590; PubMed=1569945;
 RA Demarini D.J., Winey M., Ursic D., Webb F., Culbertson M.R.;
 RT "SEN1, a positive effector of trna-splicing endonuclease in
 RT Saccharomyces cerevisiae".
 RL Mol. Cell. Biol. 12:2154-2164(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favell A., Fulton L., Gattung S., Greco T., Kirsten J., Kucaba T.,
 RA Hallsworth K., Hawkins J., Hillier L., Jier M., Johnson D.,
 RA Johnston L., Langston Y., Latreille P., Le T., Mardis E., Menezes S.,
 RA Miller N., Nan M., Pauley A., Peluso D., Rifken L., Riles L.,
 RA Taich A., Trevasakis E., Vignati D., Wilcox L., Wohlman P., Vaudin M.,
 RA Wilson R., Waterston R.;
 RL Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 CC -I- FUNCTION: REQUIRED FOR ENDONUCLYTIC CLEAVAGE OF INTRONS FROM
 CC ALL FAMILIES OF PRECURSOR TRNAS. MAY BE ONE OF SEVERAL COMPONENTS
 CC OF A NUCLEAR-LOCALIZED SPLICING COMPLEX. SEN1 IS ESSENTIAL FOR
 CC VEGETATIVE GROWTH. ESSENTIAL FUNCTIONS OF THE PROTEIN RESIDE IN
 CC THE C-TERMINAL 1214 AMINO ACIDS.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR.
 CC -I- SIMILARITY: BELONGS TO THE DNA2/NAM7 HELICASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL; M74589; AAB63976.1; -;
 CC DR EMBL; U20939; AAB67502.1; -;
 CC DR PIR: A44387; A44387.
 CC DR SGP: S0004422; SEN1.
 CC DR LRNA Processing; Nucleic protein; ATP-binding; Hydrolase; Nuclease.
 CC KM LRNA BIND 1357 1364 ATP (POTENTIAL).
 CC FT DOMAIN 1908 1961 LYS-RICH (BASIS).
 CC FT DOMAIN 1909 1927 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 CC FT VARIANT 1747 1747 G -> A (IN SEN1-1; GIVES RISE TO A
 CC TEMPERATURE SENSITIVE MUTANT).
 CC FT SEQUENCE 2231 AA; 252495 MW; 88F0FF95B3A8BD89 CRC64;

Query Match 4.5%; Score 85.5; DB 1; Length 2231;
 Best Local Similarity 18.2%; Pred. No. 26;
 Matches 66; Conservative 63; Mismatches 140; Indels 93; Gaps 15;

QY 40 PRSSKHLSPKADVKINMSYVTKTKAINCKYHREPLGRFPYLYITPMKGLQMLMAD 99
 Db 193 PHMLRLNKKQKATFEALFEKFFY-DTK-----HRLDVTNP--LSIKFISGVICWCE 242
 QY 100 AKKARRIKITNMWK-----HNKIFHQLPYREMEHLRQFQODYTKCLGIIISPPFAN 152
 Db 243 GSK-----EENEMSKRAFLKLDYSRNF-----INLSNLTPIIEEYTHILFLQNPANW 291
 QY 153 LVFL-----LMYLFPROLLIRHFWTPKQ-----QTFPDIYHAFKQS 190
 Db 292 TELVVSQFWSRLPLPVNLPFKDVEIEFYQPKVNESLAKTFKPLPEIFPMWNLHLSKV 351
 QY 191 HPEIISYLEKVIPLDISDAGLRWL---TDLCTKIQGTGTHPAIDILALRECF----- 239
 Db 352 HDKLDLFLR-----GLTMFLNKGSSEFSWKIEPFTSHLIDILFNDSPIKLIK 403
 QY 240 SNHPLGNNQOLALHVKALSRAMLITSLPPLRLRHRTKTTV--IHQDLKALAKIGQ 297
 Db 404 QDNPIVEHTEVEYFQLTGVSVDLISWLP--FYHALSPSKRIQWRYKSMAPLRIITANY 460
 QY 298 LTAQEVKASACYLGNLSTHGEDRCFTWLGEMLOISCKEAPLSLLHN----VVLIS 352
 Db 461 PSLSIKPKACILMSATL-----LRAVLTIKENRAMLYKKNDEFEYVLLTK 506
 QY 353 TN 354
 Db 507 TD 508

RESULT 7
 SYR_CHLPN STANDARD: PRT: 561 AA.
 ID SYR_CHLPN
 AC 0927Y3; 09J078;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ARGINYL-TRNA SYNTHETASE (EC 6.1.1.19) (ARGININE--TRNA LIGASE) (ARGRS).
 GN ARGS OR CPN0570 OR CP0179.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CWL029;
 RA Kelmman S., Mitchell W., Marathe R., Tamme C., Fan J., Hyman R.W.,
 RX MEDLINE=99206606; PubMed=10192388;
 RA Kelmman S., Mitchell W., Marathe R., Tamme C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis".
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Uetzerback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gwin M., Nelson W., DeBoy R., Koloney J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 RT pneumoniae AR39".
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kunara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA".
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -I- CATALYTIC ACTIVITY: AMP + L-ARGININE + TRNA(ARG) = AMP +
 CC PYROPHOSPHATE + L-ARGINYL-TRNA(ARG).
 CC -I- SUBUNIT: MONOMER (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE001641; AAD18710.1; -;
 CC EMBL: AE002179; AAF38053.1; -;
 CC EMBL: AP002547; BAA98776.1; -;
 CC TIGR: CP0179; -;
 CC InterPro: IPR001278; trna-synt_1d.
 CC InterPro: IPR001412; trna-synt_1t.
 CC Pfam: PF00750; trna-synt_1d; 1.
 CC PRINTS: PRO1038; TRNASYNTHARG.
 CC PROSITE: PS00178; AA_TRNA_LIGASE_1; 1.
 CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 CC Complete proteome. 133 "HIGH" REGION.
 CC SITE 125 366 370 "KMSKS" REGION.
 CC BINDING 369 369 ATP (BY SIMILARITY).
 CC SEQUENCE 561 AA; 63368 MW; 6386F0FF190C22DB CRC64;

Query Match 4.4%; Score 84; DB 1; Length 561;
 Best Local Similarity 23.7%; Pred. No. 6.5;
 Matches 44; Conservative 23; Mismatches 59; Indels 60; Gaps 8;

QY 178 DFLLDYHAFKQSH-----PEIISYLEK-----VPLI---SPAGLR 211
 DB 242 DLIIDIVVEKGESEYNFLEPEIIEDEKGLLVNSNDKACVFHEARSIPPMVQSGGYV 301
 QY 212 WRLTDLCTKIQRGTHPAIHDLALRECFSNHPLGMNOLAHVKALSRAMLTSYLPPL 271
 DB 302 YATFDLAMYRYIEDHAKIIIVTD-----LG---GSLHQLLEATVAIAAGYQPGI 351
 QY 272 LRH-----RLKTHP-----TVIHQLDALKALIGCOLTAQEVKSAC 308
 DB 352 FSHVGFELVLDPOGKKLTKRSGENWKLRELLDYAIKAEALKE-HRPELTDIAIDERPAP 410
 QY 309 LRLGINS 314
 DB 411 VIGINA 416
 RESULT 8
 SYM_HELPY
 ID SYM_HELPY STANDARD; PRT; 650 AA.
 AC P56127;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE METHIONYL-TRNA SYNTHETASE (EC 6.1.1.10) (METHIONINE--TRNA LIGASE)
 DE (METS).
 GN METG OR HP0417.
 OS Helicobacter pylori (Campylobacter pylori).
 OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
 OC Helicobacter.
 OX NCBI_TaxID=210;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=26695 / ATCC 700392;
 RX MEDLINE=9739467; PubMed=9252185;
 RA Tomb J.-F., White O., Keriavage A.R., Clayton R.A., Sutton G.G.,
 RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
 RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
 RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
 RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
 RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,

RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Matthey L., Wallin E.,
 RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
 RA Venter J.C.;
 RT "The complete genome sequence of the gastric pathogen Helicobacter
 RT pylori.";
 RL Nature 388:539-547(1997).
 CC -1- FUNCTION: IT IS PROBABLY ESSENTIAL FOR CELL SURVIVAL, BEING
 CC REQUIRED NOT ONLY FOR ELONGATION OF PROTEIN SYNTHESIS BUT ALSO
 CC FOR THE INITIATION OF ALL MRNA TRANSLATION THROUGH INITIATOR
 CC TRNA (PHE?) AMINOACYLATION.
 CC -1- CATALYTIC ACTIVITY: ATP + L-METHIONINE + TRNA(MET) = AMP +
 CC PYROPHOSPHATE + L-METHIONYL-TRNA(MET).
 CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC STRONG: TO CSTEINYL-TRNA SYNTHETASE.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AE000557; AAD07483.1; -;
 CC TIGR: HP0417; -;
 CC InterPro: IPR002300; trna-synt_1a.
 CC InterPro: IPR001412; trna-synt_1t.
 CC InterPro: IPR002304; trna-synt_met.
 CC InterPro: IPR002547; trna_bind.
 CC Pfam: PF00133; trna-synt_1; 1.
 CC Pfam: PF01588; trna_bind; 1.
 CC PRINTS: PRO1041; TRNASYNTHET.
 CC PROSITE: PS00178; AA_TRNA_LIGASE_1; FALSE_NEG.
 CC Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 CC trna-binding; Complete proteome.
 CC SITE 9 21 "HIGH" REGION.
 CC SITE 301 305 "KMSKS" REGION.
 CC BINDING 304 304 ATP (BY SIMILARITY).
 CC DOMAIN 555 648 TRNA BINDING.
 CC SEQUENCE 650 AA; 74506 MW; CEFDF457B9F75DD3 CRC64;

Query Match 4.4%; Score 84; DB 1; Length 650;
 Best Local Similarity 17.7%; Pred. No. 7.7;
 Matches 49; Conservative 48; Mismatches 116; Indels 64; Gaps 8;

QY 28 IQLGRSGLAWAPRSSKHLSPKADYKNLMSYVYTKTAINGKYHRTGRHPPRYIIYT 87
 DB 204 LSIPTSEFWGIPLPKKM-DEKHVVYVWLDALLNYASALG----- 243
 QY 88 IMKGIOMLMAADAKKARKIKTMKHNKIFHQLPYREMHRLROFROVDTKCLFIIISIP 147
 DB 244 -YLNDLNKMAHEECARH---VGKDLIRFAIYWP-----AFLMSLNP 284
 QY 148 PRANTLVFLMTLPFROLIRHFWT-----PKQOTDFLIYHAFRKOSHPEIISYLEKV 201
 DB 285 LF-----KOLCVHGMWTIEGVKMSKSGVNLVDQKATAMEGIELRYFLIRE 331
 QY 202 IPLIDAGLRW-LTDLCTKIQRGTHPAIHDLALRECFSNHPLGMNOLAHVKALSR 259
 DB 332 VPGQDGFPSKALIERINANLNLDGLNLRLLGKAMARKYFNHSIKSTIYAVYSKELEK 391
 QY 260 AMLL-----TSYLPPLLRHRLKTHHTVYTHQDKALAK 292
 DB 392 VHQILDNANSEVPKQMLAKALEELFNVDPLNKLIAK 428

RESULT 9
 UVRG_CHLNP
 ID UVRG_CHLNP STANDARD; PRT; 603 AA.
 AC Q926W6; Q92QA6;

DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE EXCINUCLEASE ABC SUBUNIT C.
 GN UVRC OR CPN0940 OR CP0921.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SOURCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=9206606; PubMed=10192388;
 RA Kaiman S., Mitchell W., Marathe R., Lamme C., Fan J., Hyman R.W.,
 RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
 RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
 RL Nat. Genet. 21:385-389(1999).
 RN [2]
 RP SOURCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
 RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
 RA Gaim M., Nelson W., Deboy R., Kolonay J., McClarty G., Salzberg S.L.,
 RA Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
 pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [3]
 RP SOURCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shiba T., Ishii K., Hattori M., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CML029 from USA.";
 RL Nucleic Acids Res. 28:2311-2314(2000).
 CC -1- FUNCTION: THE ABC EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 CATALYZES THE EXCISION REACTION OF UV-DAMAGED NUCLEOTIDE SEGMENTS
 PRODUCING OLIGOMERS HAVING THE MODIFIED BASE(S). ATTACHES TO THE
 UVRA-UVRB COMPLEX, DISPLACING UVRA, AND THE DAMAGED DNA STRAND IS
 NICKED ON BOTH SIDES OF THE DAMAGED SITE (BY SIMILARITY).
 CC -1- SUBUNIT: CONSISTS OF THREE SUBUNITS: UVRA, UVRB AND UVRC.
 CC -1- SUBCELLULAR LOCATION: CYTOSOL/PLASMA MEMBRANE (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE UVRC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 DR EMBL: AE001673; AAD19078.1; -
 DR EMBL: AE002251; AAF38706.1; -
 DR EMBL: AP002548; BAA9148.1; -
 DR TIGR: CP0921; -
 DR InterPro: IPR001943; UVR.
 DR InterPro: IPR001162; UVR_C_2.
 DR InterPro: IPR000305; UVR_C_1.
 DR Pfam: PF01541; Excl_endo_N_1.
 DR Pfam: PF02151; UVR_1.
 DR Pfam: PF005870; UVR_C_2; 1.
 DR ProDom: PD005870; UVR_C_2; 1.
 DR SMART: SM00465; G1YC; 1.
 DR SOS response; Excision nuclease; DNA repair; Complete proteome.
 KW SEQUENCE 603 AA; 69426 MW; 28017BAF19FEC84 CRC64;
 SQ

QY 88 IFMG-LQMLNADAKKARIKTNMKINIKFHP--YREMEHRO--FRQDYTKCLFLG 142
 Db 198 LFLKKEEYVKKLEKVIQKASD---NLEFQANRYRTSLIKQAMAKQVKEKPHQN 253
 QY 143 IISIPFANYIVFLMYLPPRO--LLIRHWPQKQOTDFLDIYAEKQ---SHP----- 192
 Db 254 IDALGLYRHKQFTLLTLTVRSGLKGARHFSFENAEODDLSSFTLQYVVSQPIYK 313
 QY 193 ELISYLEKVIPLIS-----DAGIRMLTUTL-----CTKIQKSTHP-- 227
 Db 314 ELTLPLEFPLSYVLNAESPRLRSRPTGKELDLAVRNKAYATTLPSSTLPYQ 373
 QY 228 ALHDLIALR-----ECFSN-----HPLGM-----NQLQ 250
 Db 374 DQONILRMQYRYRECYDNAMQGHATGYIVFENNPFQKQYRFSIDSEKTONDLA 433
 QY 251 ALHVALSRAMLTSLYLPPLLRHRLKTHYIVHQLKALAKLGIGULTAEVKSACLR 310
 Db 434 LLEEVLLRFHSLTALPDMIVDGGKTHYKTKKIOTLTLTGIVVTAKEKSN-HSR 492
 QY 311 GUNSTHIGEDRCRTW 325
 Db 493 GLNKERT---FCETP 504

RESULT 10
 NUCS_LIGVU STRAND: PRT; 738 AA.
 ID NUCS_LIGVU
 AC Q9TUA3;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE NADH-PLASTOQUINONE OXIDOREDUCTASE CHAIN 5, CHLOROPLAST (EC 1.6.5.3).
 GN NDHF.
 OS Ligustrum vulgare (Common privet).
 OC Chloroplast.
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 OC Asteridae; euasterids I; Lamiales; Oleaceae; Ligustrum.
 OX NCBI_TaxID=13597;
 RN [1]
 RP SOURCE FROM N.A.
 RA Kim K.-J., Jansen R.K., Olmstead R.C.;
 RT Multiple origins of sympetaly and associated floral characters.";
 RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: NADH + PLASTOQUINONE = NAD(+) + PLASTOQUINOL.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@sib-sib.ch).
 CC
 CC EMBL: AFI30164; AAF08126.1; -
 DR EMBL: AFI30164; AAF08126.1; -
 DR InterPro: IPR001750; Oxidored_q1.
 DR InterPro: IPR002128; Oxidored_q1_C.
 DR InterPro: IPR001516; Oxidored_q1_N.
 DR Pfam: PF00361; Oxidored_q1; 1.
 DR Pfam: PF01010; Oxidored_q1_C; 1.
 DR Pfam: PF00662; Oxidored_q1_N; 1.
 DR Oxidoreductase; NAD; Plastocyanine; Chloroplast.
 KW SEQUENCE 738 AA; 83477 MW; CD78BCABDC1BCAF CRC64;
 SQ

Query Match 4.4%; Score 83.5; DB 1; Length 738;
 Best Local Similarity 22.3%; Pred. No. 10;
 Matches 39; Conservative 24; Mismatches 61; Indels 51; Gaps 7;

QY 59 YVVTKATNGKYNHFLGRRHPPRYLYITPMKGLQMLM--ADAKKA---RIKTNMK 112
 Db 450 YLTPFEGHIANHFQYVSKQKMTPFYSI-----SLWKGDSKRIKRNRLKMNNSK 500

```

QY 113 HNRKQHPYREMEHLRQ-----FRQDYTK-----CLFLG 142
DB 501 SSSFFSKPYRSGKNVNRGPELTIIVHFNQKSYSPRESQNTMLPPLVLGILFLFVS 560
QY 143 IISPPANVLVFLMLFPRQLLIRHFWTPKQOTDFLDIYHAFKROSHPELISY 197
DB 561 SLGI-FPNOELDLITLWLPSPINLHOKW-----NDSIDWYEPWKDASFSVSIAV 609

RESULT 11
DPOE_SCHPO STANDARD; PRT; 2199 AA.
AC P87154;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA POLYMERASE EPSILON, CATALYTIC SUBUNIT A (EC 2.7.7.7) (DNA
DE POLYMERASE II SUBUNIT A).
GN CDC20 OR SPBC25H2.13C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=972;
RA Duroso G., Lye G., Bowman S., Church C., Wood V., Barrell B.G.,
RA Rajandream M.A., Connor R.E.;
RA Submitted (May-1997) to the EMBL/Genbank/DBJ databases.
CC -! FUNCTION: DNA POLYMERASE II PARTICIPATES IN CHROMOSOMAL DNA
CC REPLICATION (BY SIMILARITY).
CC -! CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE =
CC N PYROPHOSPHATE + DNA(N).
CC -! SUBUNIT: CONSISTS OF FIVE SUBUNITS (200 KDA, 80 KDA, 34 KDA, 30
CC KDA, AND 29 KDA) (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -! DOMAIN: THE DNA POLYMERASE ACTIVITY DOMAIN RESIDES IN THE
CC N-TERMINAL HALF OF THE PROTEIN, WHILE THE C-TERMINUS IS NECESSARY
CC FOR COMPLEXING SUBUNITS B AND C (BY SIMILARITY).
CC -! MISCELLANEOUS: IN EUKARYOTES THERE ARE FIVE DNA POLYMERASES:
CC ALPHA, BETA, GAMMA, DELTA, AND EPSILON WHICH ARE RESPONSIBLE FOR
CC DIFFERENT REACTIONS OF DNA SYNTHESIS.
CC -! SIMILARITY: BELONGS TO DNA POLYMERASE TYPE-B FAMILY. HIGH
CC SIMILARITY WITH MAMMALIAN DNA POLYMERASE EPSILON.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z93397; CAB08772.1; -
DR InterPro; IPR002064; DNA-pol.B.
DR Pfam; PF00136; DNA.pol.B. 2.
DR PRINTS; PR00106; DNAPOLB.
DR SMART; SM00486; POLBC. 1.
DR TRANSFERASE; DNA-directed DNA polymerase; DNA replication;
KW DNA-binding; Zinc-finger; Nuclear protein.
FT ZN_FING 2069 2155 BY SIMILARITY.
FT SEQUENCE 2199 AA; 252885 MW; A80A5D0865EEBC3E CMC64;

Query Match 4.38; Score 82.5; DB 1; Length 2199;
Best Local Similarity 26.7%; Pred. NO. 47;
Matches 36; Conservative 14; Mismatches 46; Indels 39; Gaps 5;

QY 177 TDFDIDYHAFKQSHPELISYLEKVIPLISDAGLRMLTLGLCKRIQGTGHPALHDLALR 236
DB 462 SDAAVATFELMKYVHPITFS-LCNILPLNDEVLK-----KGVTGLCELTLLTVE 509

```

```

QY 237 EC-----FSNHPJGMNOALHVKALS-----RAMLLTSYLPPLLRHR 275
DB 510 ACTKNILLPNKHVDASOKFFEDHLLASETYVGHVESLESQVFRSDLDPTNMDP----- 564
QY 276 LKTHTVIHLQDKAL 290
DB 565 -KVEFELLQDLKAL 578

RESULT 12
AD15_RAT STANDARD; PRT; 816 AA.
ID AD15_RAT
AC Q90YV0;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ADAM 15 PRECURSOR (EC 3.4.24.-) (A DISINTEGRIN AND METALLOPROTEINASE
DE DOMAIN 15) (METALLOPROTEINASE-LIKE, DISINTEGRIN-LIKE, AND CYSTEINE-
DE RICH PROTEIN 15) (MDC-15) (METALLOPROTEINASE RGD DISINTEGRIN PROTEIN)
DE (METARGIDIN) (CRIL-7).
GN ADAM15 OR MDC15.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE=Sciatic nerve;
RA MEDLINE=20556183; PubMed=11102971;
RA Bosse F., Petzold G., Greiner-Petter R., Pippl U., Gyllen C.,
RA Mueller H.-W.;
RA "Cellular localization of the disintegrin CRIL-7/MDC15 mRNA in rat
RA PNS and CNS and regulated expression in postnatal development and
RA after nerve injury."
RA Glia 32:313-327(2000).
CC -! FUNCTION: MAY BE INVOLVED IN CELL-SURFACE PROTEOLYSIS, CELL
CC ADHESION OR INTRACELLULAR PROTEIN MATURATION.
CC -! COFACTOR: BINDS ONE ZINC ION (BY SIMILARITY).
CC -! SUBUNIT: INTERACTS WITH INTEGRIN ALPHA-V-BETA3, ENDOPHLIN I AND
CC SORTING NEXIN 9.
CC BIND THE PRECURSOR BUT NOT THE PROCESSED FORM OF ADAM15,
CC SUGGESTING THAT THE INTERACTION OCCURS IN A SECRETORY PATHWAY
CC COMPARTMENT PRIOR TO THE MEDIAL GOLGI (BY SIMILARITY).
CC -! SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -! TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN, SPINAL CORD,
CC SCIATIC NERVE AND LUNG. EXPRESSED AT LOWER LEVELS IN ALL OTHERS
CC TISSUES. IN THE PERIPHERAL NERVOUS SYSTEM, EXPRESSED PREDOMINANTLY
CC BY SCHWANN CELLS. IN THE CENTRAL NERVOUS SYSTEM, PREFERENTIALLY
CC EXPRESSED BY NEURONAL CELLS.
CC -! INDUCTION: IN RESPONSE TO SCIATIC NERVE INJURY.
CC -! DOMAIN: THE CYTOPLASMIC DOMAIN INTERACTS WITH ENDOPHLIN I AND
CC SORTING NEXIN 9 (BY SIMILARITY).
CC -! DOMAIN: DESINTEGRIN DOMAIN BINDS TO INTEGRIN ALPHA-V-BETA3 (BY
CC SIMILARITY).
CC -! PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY M12B (ZINC
CC METALLOPROTEASE); ALSO KNOWN AS THE REPOUSLIN SUBFAMILY.
CC -! SIMILARITY: CONTAINS 1 EGF-LIKE DOMAIN.
CC -! SIMILARITY: CONTAINS 1 DISINTEGRIN DOMAIN.
CC -! SIMILARITY: CONTAINS 2 SH3-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ251198; CAB61762.1; -
DR HSPD; P17494; IKST.

```

```

DR InterPro: IPR000130; Zn_Mtpeptidase.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001590; Reprolysin.
DR InterPro: IPR001762; Disintegrin.
DR InterPro: IPR002870; Pcp_M12B_propep.
DR Pfam: PF00200; disintegrin_1.
DR Pfam: PF01562; Pcp_M12B_propep_1.
DR ProDom: PD000664; Disintegrin_1.
DR SMART: SM00050; Disintegrin_1.
DR SMART: SM00181; EGF_1.
DR PROSITE: PS00215; ADAM_MERO_1.
DR PROSITE: PS00427; DISINTEGRIN_1; FALSE_NEG.
DR PROSITE: PS00214; DISINTEGRIN_2_1.
DR PROSITE: PS00022; EGF_1; FALSE_NEG.
DR PROSITE: PS01186; EGF_2_1.
DR PROSITE: PS00142; ZINC_PROTEASE_1.
DR PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
KM Hydrolase; Metalloprotease; zinc; Signal; Glycoprotein; Zymogen;
transmembrane; EGF-like domain; SH3-binding.
FT SIGNAL 1 17
FT PROPEP 18 208
FT CHAIN 209 816
FT DOMAIN 209 698
FT TRAMEM 699 719
FT DOMAIN 720 816
FT DOMAIN 209 416
FT DOMAIN 423 510
FT DOMAIN 511 658
FT DOMAIN 659 687
FT DOMAIN 700 713
FT DOMAIN 768 774
FT SITE 803 809
FT SITE 180 180
FT METAL 350 350
FT ACT_SITE 351 351
FT METAL 354 354
FT METAL 360 360
FT DISULFID 325 411
FT DISULFID 482 495
FT DISULFID 659 669
FT DISULFID 663 675
FT DISULFID 677 686
FT CARBOHYD 57 57
FT CARBOHYD 239 239
FT CARBOHYD 391 391
FT CARBOHYD 394 394
FT CARBOHYD 608 608
FT CARBOHYD 613 613
SQ SEQUENCE 816 AA; 88051 MW; B9D2CE023266FC27 CRC64;

```

Query Match 4.3%; Score 81.5; DB 1; Length 816;
 Best Local Similarity 22.7%; Pred. No. 17;
 Matches 59; Conservative 33; Mismatches 71; Indels 97; Gaps 16;

```

QY 124 EMEHLRQFRQDY--TKCLFLGIIIS-----IPFANVL-----VFLMWLPROLIR 168
DB 199 EQHNLRRLKRDVYETKIVELYADVNSEVRYKRPQQLNKRTLEVALLDLTFQPLNVR 258
QY 169 -----HWTPKQQTDFLDIYHAFKQSHPEIISTYLEKVIPLISDAGLRWRLTDLCTRIQ 222
DB 259 VALVGLQWMT---QRDLTEIEM-----SSNPAY-----LDNFLRWRTDLDLPRLP 299
QY 223 RGTHTAIDHI--LALREGEFNNHPLGMNQLQALHVKALSRAMLITSYLPPILRHR--LK 277
DB 300 -----HDSAOLVTYVTSFGPMWGM-----AIDNSICSPFSGGVNMDHSTSLIG 343
QY 278 THTTVIHLQKALAKIGI-----GOLTAQOEY-----KSACYLRGLNSTHIGEDRC 322
DB 344 VASSTIAHELGHG--LGLDHDSPGNSCPGPRAPAKSKCIMEASTDFLPLGLNFSN-----C 395
QY 323 RTW-----LGEW 329

```

```

DB 396 SRMALEKALLDGMSCLEFW 415
RESULT 13
ID RPKM_HUMAN STANDARD; PRT; 1230 AA.
AC 000411; 060370;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE, MITOCHONDRIAL PRECURSOR (EC 2.7.7.6)
DE (MTRPOL).
GN POLRMT.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver.
RX MEDLINE=97252399; PubMed=9097968;
RA Tiranti V., Savola A., Forti F., D'Apolito M.F., Centra M., Rocchi M.,
RA Zeviani M.;
RT "Identification of the gene encoding the human mitochondrial RNA
RT polymerase (h-mRPol) by cyberscreening of the Expressed Sequence
RL Hum. Mol. Genet. 6:615-625(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCready P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Barnes J., Danganan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Altix C., Andeise T.,
RA Trankheim M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Krommiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,
RA Carraro A.V.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -!- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE = N PYROPHOSPHATE +
CC RNA(N).
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -!- SIMILARITY: BELONGS TO THE PHAGE AND MITOCHONDRIAL RNA POLYMERASES
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U75370; AAB58255.1; -.
DR EMBL: AC004449; AAC06147.1; -.
DR MIM: 601778; -.
DR InterPro: IPR002092; RNA_pol_phage.
DR Pfam: PF00940; RNA_pol_1.
DR PROSITE: PS00490; RNA_POL_PHAGE_1; 1.
DR PROSITE: PS00499; RNA_POL_PHAGE_2; 1.
KW Transferase; DNA-directed RNA polymerase; Transcription;
KW Mitochondrion; Transist peptide.
FT TRANSIT 1 41
FT CHAIN 42 1230
FT ACT_SITE 922 922
FT ACT_SITE 991 991
FT ACT_SITE 1151 1151
FT CONFLICT 399 399
FT CONFLICT 983 983
SQ SEQUENCE 1230 AA; 138684 MW; 412E124D517FEC1A CRC64;

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:09:07 ; Search time 12.62 Seconds
(without alignments)
641.932 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912
Sequence: 1 MALSRVCARSAVMSAVTP.....LSLLHNVLSTNYIGTRR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapect 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	6.5	586	2	US-08-630-822A-70
2	124	6.5	586	2	US-09-005-069-70
3	94	4.9	428	4	US-09-139-064-2
4	94	4.9	428	4	US-09-139-064-4
5	94	4.9	428	4	US-09-487-370-2
6	94	4.9	428	4	US-09-487-370-4
7	87	4.6	648	1	US-08-451-715A-4
8	84	4.4	1257	2	US-08-750-152A-2
9	80.5	4.2	147	3	US-08-946-329A-58
10	80	4.2	200	1	US-07-959-284-8
11	80	4.2	200	5	US-08-308-736A-8
12	80	4.2	200	5	PCT-US93-09649A-8
13	80	4.2	200	5	PCT-US93-09649-8
14	79	4.1	452	3	US-09-052-778-15
15	79	4.1	877	3	US-08-911-853-31
16	79	4.1	877	4	US-09-479-409-31
17	78.5	4.1	451	4	US-09-357-251-35
18	78.5	4.1	2629	2	US-08-751-189-4
19	78.5	4.1	2629	2	US-09-060-836-4
20	78.5	4.1	2629	4	US-09-184-445-4
21	77	4.0	402	1	US-08-036-210-15
22	77	4.0	402	2	US-08-449-609-15
23	77	4.0	898	1	US-08-036-210-22
24	77	4.0	898	2	US-08-449-609-22
25	76.5	4.0	500	2	US-08-578-709-15
26	76.5	4.0	1151	3	US-08-840-006-6
27	76.5	4.0	1200	3	US-08-840-006-5

28	75.5	3.9	472	4	US-08-976-255-17	Sequence 17, Appl
29	75.5	3.9	1005	1	US-08-089-986-3	Sequence 3, Appl
30	75.5	3.9	1005	1	US-08-478-585-3	Sequence 3, Appl
31	75.5	3.9	1005	1	US-08-717-312-3	Sequence 3, Appl
32	75.5	3.9	1005	2	US-08-266-408-3	Sequence 3, Appl
33	75.5	3.9	1005	5	PCT-US94-07886-3	Sequence 3, Appl
34	75	3.9	870	1	US-08-190-687B-25	Sequence 25, Appl
35	75	3.9	870	5	PCT-US94-00198-5	Sequence 5, Appl
36	75	3.9	1047	1	US-08-190-687B-8	Sequence 8, Appl
37	74.5	3.9	336	1	US-08-176-427B-6	Sequence 6, Appl
38	74.5	3.9	336	2	US-08-356-060A-10	Sequence 10, Appl
39	74.5	3.9	336	4	US-08-674-509B-10	Sequence 10, Appl
40	74.5	3.9	411	4	US-08-460-900C-10	Sequence 10, Appl
41	74.5	3.9	411	4	US-08-954-698-10	Sequence 4, Appl
42	74.5	3.9	492	1	US-07-783-705A-4	Sequence 4, Appl
43	74	3.9	326	2	US-09-055-097-6	Sequence 6, Appl
44	73.5	3.8	1064	3	US-08-726-214-8	Sequence 8, Appl
45	73	3.8	614	1	US-08-543-881-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-630-822A-70
; Sequence 70, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
; APPLICANT: FRANK, GLENN R.
; APPLICANT: HUNTER, SHIRLEY WU
; APPLICANT: WALLENFELS, LYNDIA
; TITLE OF INVENTION: NOVEL ECTOPARASITIC SALIVA PROTEINS
; TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
; NUMBER OF SEQUENCES: 107
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,822A
; FILING DATE: 11-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CONNELL, GARY J.
; REGISTRATION NUMBER: 32,020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 586 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Xaa = any amino acid
; LOCATION: 379
; US-08-630-822A-70

Query Match 6.5%; Score 124; DB 2; Length 586;
Best local Similarity 27.0%; Pred. No. 3.3e-05;
Matches 37; Conservative 29; Mismatches 69; Indels 2; Gaps 2;

APPLICANT: Biswas, Sanjoy
TITLE OF INVENTION: Response Regulator
FILE REFERENCE: GM10091
CURRENT APPLICATION NUMBER: US/09/139,064
CURRENT FILING DATE: 1998-08-25
EARLIER APPLICATION NUMBER: 60/060,714
EARLIER FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 428
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-139-064-4

Query Match 4.9%; Score 94; DB 4; Length 428;
Best Local Similarity 20.8%; Pred. No. 0.047;
Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

OY 29 QLRSGGLAWGAPRSSKHLSPKADVKNLMSYVYTKAKINCKYHRLGHPREYILYTI 88
DB 122 QLGERG-----KKSQTLSELDGAFVSYLDKENMMWIGLSKEKQ-----STTIPYIV 170
OY 89 FMKGLQMLADAKKARRIKTNMMKNHFKHQLPYREMEHLROFRODVTKCLFLGIISIP 148
DB 171 LGQAMQIFISDQPLDGLVYTF-----EAPYQ--EHFERKWLNAEKTLEFGSVNLQ 220
OY 149 ----FA-----NYLVFLMLTFPR-----QLLR-----HFWT 172
DB 221 SPSLFAVYEPYRVIIQGNLQIYEELNLEKVLLENTPRIPITKOLFQFVMDVPHLE 280
OY 173 PROQDFLDI---YHAFKQSHPEIISYLEKVIPLSDAGLRRLDICTKIQGTNPAT 229
DB 281 HKADMDIVKTHAI--QSFDELVSYKET--LISFEG-QYRMNE-----322
OY 230 HDLALRECFSNHPLGMQOLAHVKALSRMLLTSLPPLLRHRLKTHVYIHOIDKA 289
DB 323 -NVSVLEV-----IGRDYKELSLDKSKALFINPYVLGQLIKR--ETDSTFAELLNKO 374
OY 290 LAKLIGOL--TAQEVKSACYLRGLNS 314
DB 375 RIRAAQOLLSTSDSTENICVAVGYSN 401

RESULT 5
US-09-487-370-2
Sequence 2, Application US/09487370
Patent No. 6294652

GENERAL INFORMATION:
APPLICANT: Wallis, Nicola G.
APPLICANT: Ingraham, Karen A.
APPLICANT: Ge, Yigong
APPLICANT: Holmes, David J.
APPLICANT: Zalacain, Magdalena
APPLICANT: Throup, John
APPLICANT: Biswas, Sanjoy
TITLE OF INVENTION: Response Regulator
FILE REFERENCE: GM10091
CURRENT APPLICATION NUMBER: US/09/487,370
CURRENT FILING DATE: 2000-01-18
PRIOR APPLICATION NUMBER: 09/139,064
PRIOR FILING DATE: 1998-08-25
PRIOR APPLICATION NUMBER: 60/060,714
PRIOR FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 428
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-487-370-2

Query Match 4.9%; Score 94; DB 4; Length 428;
Best Local Similarity 20.8%; Pred. No. 0.047;
Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

OY 29 QLRSGGLAWGAPRSSKHLSPKADVKNLMSYVYTKAKINCKYHRLGHPREYILYTI 88
DB 122 QLGERG-----KKSQTLSELDGAFVSYLDKENMMWIGLSKEKQ-----STTIPYIV 170
OY 89 FMKGLQMLADAKKARRIKTNMMKNHFKHQLPYREMEHLROFRODVTKCLFLGIISIP 148
DB 171 LGQAMQIFISDQPLDGLVYTF-----EAPYQ--EHFERKWLNAEKTLEFGSVNLQ 220
OY 149 ----FA-----NYLVFLMLTFPR-----QLLR-----HFWT 172
DB 221 SPSLFAVYEPYRVIIQGNLQIYEELNLEKVLLENTPRIPITKOLFQFVMDVPHLE 280
OY 173 PROQDFLDI---YHAFKQSHPEIISYLEKVIPLSDAGLRRLDICTKIQGTNPAT 229
DB 281 HKADMDIVKTHAI--QSFDELVSYKET--LISFEG-QYRMNE-----322
OY 230 HDLALRECFSNHPLGMQOLAHVKALSRMLLTSLPPLLRHRLKTHVYIHOIDKA 289
DB 323 -NVSVLEV-----IGRDYKELSLDKSKALFINPYVLGQLIKR--ETDSTFAELLNKO 374
OY 290 LAKLIGOL--TAQEVKSACYLRGLNS 314
DB 375 RIRAAQOLLSTSDSTENICVAVGYSN 401

RESULT 6
US-09-487-370-4
Sequence 4, Application US/09487370
Patent No. 6294652

GENERAL INFORMATION:
APPLICANT: Wallis, Nicola G.
APPLICANT: Ingraham, Karen A.
APPLICANT: Ge, Yigong
APPLICANT: Holmes, David J.
APPLICANT: Zalacain, Magdalena
APPLICANT: Throup, John
APPLICANT: Biswas, Sanjoy
TITLE OF INVENTION: Response Regulator
FILE REFERENCE: GM10091
CURRENT APPLICATION NUMBER: US/09/487,370
CURRENT FILING DATE: 2000-01-18
PRIOR APPLICATION NUMBER: 09/139,064
PRIOR FILING DATE: 1998-08-25
PRIOR APPLICATION NUMBER: 60/060,714
PRIOR FILING DATE: 1997-09-09
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 428
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-487-370-4

Query Match 4.9%; Score 94; DB 4; Length 428;
Best Local Similarity 20.8%; Pred. No. 0.047;
Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

OY 29 QLRSGGLAWGAPRSSKHLSPKADVKNLMSYVYTKAKINCKYHRLGHPREYILYTI 88
DB 122 QLGERG-----KKSQTLSELDGAFVSYLDKENMMWIGLSKEKQ-----STTIPYIV 170
OY 89 FMKGLQMLADAKKARRIKTNMMKNHFKHQLPYREMEHLROFRODVTKCLFLGIISIP 148
DB 171 LGQAMQIFISDQPLDGLVYTF-----EAPYQ--EHFERKWLNAEKTLEFGSVNLQ 220
OY 149 ----FA-----NYLVFLMLTFPR-----QLLR-----HFWT 172

```

RESULT 7
US-08-451-715A-4
: Sequence 4, Application US/08451715A
: Patent No. 5801013
: GENERAL INFORMATION:
: APPLICANT: Tao, Jianshi
: APPLICANT: Qiu, Yan
: APPLICANT: Houman, Fariba
: APPLICANT: Shen, Xiaoyu
: APPLICANT: Schimmel, Paul R.
: TITLE OF INVENTION: Helicobacter Aminoacyl-tRNA Synthetase
: TITLE OF INVENTION: Proteins, Nucleic Acids and Strains Comprising Same
: NUMBER OF SEQUENCES: 67
: CORRESPONDENCE ADDRESSES:
: ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
: STREET: Two Millitia Drive
: CITY: Lexington
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02173
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/451,715A
: FILING DATE: 26-MAY-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Brook, David E.
: REGISTRATION NUMBER: 22,592
: REFERENCE/DOCKET NUMBER: CPI94-25
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-861-6240
: TELEFAX: 617-861-9540
: INFORMATION FOR SPO ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 648 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-451-715A-4

```

Query Match	Best Local Similarity	4.6%;	Score 87;	DB 1;	Length 648;
Matches	50;	Conservative	49;	Mismatches	114;
				Indels	64;
				Gaps	
QY	28	LGTRGSLMGAPRSSKSLSPKADVUKMLMYYVTKTKALNGYHFFGLRHPREFILYLT	87		
Db	204	LSIRTFSEWGPJLPKKN-DPKHVVYVWLALNLNASALG-----	245		
QY	88	IFMKGIOMLAQAKKARKIKTNMKMHNIKFHODPYREMEHROFYRQDYTKCLFGLGISIP	147		
Db	244	-YINGDKNMAHFECARI--VGKDLIRFHAIYMP-----AFLMSTLIP	284		
QY	148	PFANYLVFLMLTFPROLLIRHFWT-----PQGDYDFLDIYNHARKKSHSIIYSLEKY	201		

```

Db      285  LF-----KQLCVHOMWTIEGVKSKSLGVNDAOKLAMEVGIETLRYFLHRE 331
Oy      202  IPLLSDAGLWR--LTDLCTKIQGTHPAIDIALIECEPNSPIGLMOLOLHYKALS 255
Db      332  VPPGODDESKKALIERIANLNDGLNLNRLGLMAKRYFNLSKSKITAIYSELEK 391
Oy      260  AMLL-----TSYLPPLLRRLRKTHTYVHQDLDAALK 292
Db      392  AHQILDNANSFVPMQMLKALEELFNVIDLNLKLIK 428

```

```

1      RESULT      8
2      US-08-750-152A-2
3      : Sequence 2, Application US/08750152A
4      : Patent No. 597731
5      :
6      : GENERAL INFORMATION:
7      :
8      : APPLICANT: ASAKURA, YOKO
9      : APPLICANT: KIMURA, EIICHIRO
10     : APPLICANT: ABE, CHIZU
11     : APPLICANT: KAWAHARA, YOSHIO
12     : APPLICANT: NAKAMATSU, TATSUYOSHI
13     : TITLE OF INVENTION: ALPHA-KETOGLUTARATE DEHYDROGENASE GENE
14     : NUMBER OF SEQUENCES: 14
15     :
16     : CORRESPONDENCE ADDRESSES:
17     :
18     : ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
19     : STREET: 1735 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
20     : CITY: ARLINGTON
21     : STATE: VA
22     :
23     : COUNTRY: USA
24     :
25     : ZIP: 22152
26     :
27     : COMPUTER READABLE FORM:
28     :
29     : MEDIUM TYPE: floppy disk
30     : COMPUTER: IBM PC compatible
31     : OPERATING SYSTEM: PC-DOS/MS-DOS
32     : SOFTWARE: FASTSEQ Version 1.5
33     :
34     : CURRENT APPLICATION DATA:
35     :
36     : APPLICATION NUMBER: US/08/750,152A
37     :
38     : FILING DATE:
39     :
40     : CLASSIFICATION: 435
41     :
42     : PRIOR APPLICATION DATA:
43     :
44     : APPLICATION NUMBER:
45     :
46     : FILING DATE:
47     :
48     : ATTORNEY/AGENT INFORMATION:
49     :
50     : NAME: NORMAN F. OBLON
51     : REGISTRATION NUMBER: 24,618
52     : TELECOMMUNICATION INFORMATION:
53     :
54     : TELEPHONE: 703-413-3000
55     : TELEFAX: 703-413-2220
56     :
57     : INFORMATION FOR SEQ ID NO: 2:
58     :
59     : SEQUENCE CHARACTERISTICS:
60     :
61     : LENGTH: 1257 amino acids
62     : TYPE: amino acid
63     :
64     : TOPOLOGY: linear
65     :
66     : MOLECULE TYPE: protein
67     :
68     : US-08-750-152A-2

```

Query Match	4.4%	Score 84	DB 2	Length 1257
Best Local Similarity	29.5%	Pred. No. 3.2		
Matches 38	Conservative 20	Mismatches 45	Indels 26	Gaps 8

QY	258	SRAMLL-----SYLPP--PLRLNR---LKTHTTVIHOLDALAKLIG--OUTAOEYK	304
		: :	
Db	417	SRGHLIDTNPISWVPGMVPDHRDLDIETHSILTIMDLDTFVSFGSGKREMTITREVL	476
QY	305	S---ACYLRGLNS--THIGEDRCRTMIGEMQLQISC-SIKAEISLLAHNY-----VLL	351
		: :	
Db	477	SRRLAAITLVKGESEYTHILDRDEFTWLODRLEAGMPKPTQAEQYIILQKNAAEFENFL	536
QY	352	STNYLGTTR	360
		: : :	
Db	537	QTRVYGOKR	545

INFORMATION FOR SEQ ID NO: 8:

Accession	Conservative	Mismatches	Indels	Gaps
97 WADAKARIKTNMKHNKFFHQLPYREME---HLRQFRODVTKCLEGLISIPPEANY 152	34;	61;	44;	8;

```

D6  MSELTEARLQDGNQAVR-TEHVLLARLSDQGNPTREGDFHQALNHLLDQAAGA-- 120
OY  LVFLMLTFRRQOLLRNHWPRKQOTDRLDYLNARKQSNHEPITSYLEKVLPLSDAGL-- 210
D6  -----VOIEELMLLEKKIPRNADMPNR-----VGDGGLFE 153
OY  153 LVFLMLTFRRQOLLRNHWPRKQOTDRLDYLNARKQSNHEPITSYLEKVLPLSDAGL-- 210
D6  121 -----VOIEELMLLEKKIPRNADMPNR-----VGDGGLFE 153
OY  211 --RWRLDYLCIORGTNPRAINDLALRECSNPHRLGMNOLALNVALSRAM 261
D6  154 KKLGLGKLVL-OELSQWVYRSHIDLRLV-----SSHQMGISALEL-HYAKDKQM 200

```

```

RESULT 14
US-09-052-778-15
: Sequence 15, Application US/09052778A
: Patent No. 6060590
: GENERAL INFORMATION:
: APPLICANT: Bryan, Peter J.
: APPLICANT: Kawamura, Kazuo
: TITLE OF INVENTION: CHITINASE RELATED PROT
: TITLE OF INVENTION: OF USE
: FILE REFERENCE: 07306/015001
: CURRENT APPLICATION NUMBER: US/09/052,778A
: CURRENT FILING DATE: 1996-03-31
: NUMBER OF SEQ ID NOS: 16
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 15
: LENGTH: 452
: type: prot
: ORGANISM: Drosophila melanogaster
US-09-052-778-15

```

Query Match	4.1%;	Score 79;	DB 3;	Length 452;
Best Local Similarity	22.7%;	Pred. No. 2.5;		
Matches	51;	Conservative	27;	Mismatches 75; Indels 72; Gaps 13

```

QY 52 DYKNLMSTVYTK-----TRAINCKYHFRGHRFPRIYLYTF-----MGLOMLM 97
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 NKILLISVGGDKDIEDKDAELPKYTLLELSEPTGRFRFNVTYSLVITYGFDGLVAM 164
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 98 ADAC-KARRIKT--NMKKNHNIKFHQ---LPRYEMELHRO-----FROD----- 134
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 QEPKKKPKKVVHSGIONLMKGGFKVYSSGSIDYDEKSEEHKEBFOTALRVKNAAFRPNLL 224
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 135 -----YTKLEFGIISIPRPANYLVFLMYLFRPOLLIRHEWTPKOQDFFLD---IY 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 STTVLPNNSSLEF---YDIPAVNYLDEVNLTGF-----DEFTPOGNNPEIADYAAPIY 274
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 184 HAFKROSHPEI-----ISY-----LEKVPILSDAQLRWRDLD 216
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 275 EL--SERDEENVAOVAQYWLERNNCPSAKINIVGAITYGRPKLLD 317

```

RESULT 15
US-08-911-853-31
; Sequence 31, Application US/08911853

```

1  GENERAL INFORMATION:
2
3  APPLICANT: Gerlisse, Gijlsbert
4
5  APPLICANT: Quax, Wilhelmus J.
6
7  TITLE OF INVENTION: EXPRESSION SYSTEM FOR ALTERED
8
9  TITLE OF INVENTION: EXPRESSION LEVELS
10
11 NUMBER OF SEQUENCES: 37
12
13 CORRESPONDENCE ADDRESS:
14
15 ADDRESS: Genencor International
16
17 STREET: 925 Page Mill Road
18
19 CITY: Palo Alto
20
21 STATE: CA
22
23 COUNTRY: USA
24
25 ZIP: 94304-1013
26
27 COMPUTER READABLE FORM:
28
29 MEDIUM type: Diskette
30

```

```

1  COMPUTER: IBM Compatible
2  OPERATING SYSTEM: DOS
3  SOFTWARE: FastSeq for Windows Version 2.0
4  CURRENT APPLICATION DATA:
5  APPLICATION NUMBER: US/08/911,853
6  FILING DATE:
7  PRIOR APPLICATION DATA:
8  APPLICATION NUMBER: 08/699,092
9  FILING DATE: 16-AUG-1996
10 ATTORNEY/AGENT INFORMATION:
11 NAME: Glaister, Debra J
12 REGISTRATION NUMBER: 33,888
13 REFERENCE/DOCKET NUMBER: GC361-2
14 TELECOMMUNICATION INFORMATION:
15 TELEPHONE: 650-846-7620
16 TELEFAX: 650-845-6504
17 INFORMATION FOR SEQ ID NO: 31:
18 SEQUENCE CHARACTERISTICS:
19 LENGTH: 877 amino acids
20 TYPE: amino acid
21 STRANDEDNESS: single
22 TOPOLOGY: linear
23
24 US-08-911-853-31

```

Query Match	4.1%;	Score 79;	DB 3;	Length 877;
Best Local Similarity	20.3%;	Pred. No. 6.7;		
Matches	86;	Conservative	63;	Mismatches 152;
				Indels 122;
				Gaps 22.

```

0Y      8  WARSJWMSAV-----TPGHFVTR-----RLQGRSLGAGARSSKILHSPK---50
Db      49  WQASRPFSVVHYHPLQARDNRPVAFRRRLLEASTAQVEDDELSPFNFAEMHQAPVIG 108
0Y      51  --ADVK-----LMSVVTKKAINGKYHPL-GRHPREYILY 86
Db      109  EYALADNRIESRLYLVDQCIGQPIILDYLSAMERLALGNRTVLLSRNMGF----164
0Y      87  TTFMGLOMLNADKARKRITNNMKHNKRNOLRYEMHELROFQRODYKCLFGLIISI 146
Db      165  ----SLSRKIDKN-----LICIDQHMRLSPV---QIDHNLHY-----LGG 199
0Y      147  PPFANYLFLIMLYEP-----ROLLRNHEWTPKQOTDELIVAFPKOSHPELIISY--197
Db      200  PELSAYVYGLKAMTEGMYGVKALMHN---AREGTEALQRFEG-----GHPEIVDYFGH 252
0Y      198  --LEKVITPLISDAGIKRWL-----TDLCTKI-QGSTRNAINDILALRECF-----SNRPFG 245
Db      253  VVLKRLSPQLHDFELCSAIFERFDELDRLVDSGSAILEDIAELEFLPVDEPR-G 311
0Y      246  MNQOLALHYKALSRAMLTSYLPRLPLRNRYKRTQTVVHOLDKALAKIGLOLPAQEVKS 305
Db      312  CYRYHA-----LHDFLARLAVHKQEVAAQDLHRRALALDQQRGDELALQNAQR 361
0Y      306  ACYLRGLNSTHIGEDRCRTWLG-----EWLOISCSLKFAEELSLLHNVLSTWYLG 357
Db      362  SGD-PALEQSMIGE-ACQOWYRSGHFAEVLKMLE---PLSAEELXOSRRLVLVMTYATL 416
0Y      358  TRR 360
Db      417  SRR 419

```

Search completed: December 20, 2001, 10:11:05
Job time: 118 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:09:32 (without alignments)
1070.939 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912
Sequence: 1 MALSRYCWARSAMGSAVTP.....LSLLHNVLSTINYLGR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1:	/SIDS8/gcgcata/gemseq/gemseqp/AA1980.DAT *
2:	/SIDS8/gcgcata/gemseq/gemseqp/AA1981.DAT *
3:	/SIDS8/gcgcata/gemseq/gemseqp/AA1982.DAT *
4:	/SIDS8/gcgcata/gemseq/gemseqp/AA1983.DAT *
5:	/SIDS8/gcgcata/gemseq/gemseqp/AA1984.DAT *
6:	/SIDS8/gcgcata/gemseq/gemseqp/AA1985.DAT *
7:	/SIDS8/gcgcata/gemseq/gemseqp/AA1986.DAT *
8:	/SIDS8/gcgcata/gemseq/gemseqp/AA1987.DAT *
9:	/SIDS8/gcgcata/gemseq/gemseqp/AA1988.DAT *
10:	/SIDS8/gcgcata/gemseq/gemseqp/AA1990.DAT *
11:	/SIDS8/gcgcata/gemseq/gemseqp/AA1991.DAT *
12:	/SIDS8/gcgcata/gemseq/gemseqp/AA1992.DAT *
13:	/SIDS8/gcgcata/gemseq/gemseqp/AA1993.DAT *
14:	/SIDS8/gcgcata/gemseq/gemseqp/AA1994.DAT *
15:	/SIDS8/gcgcata/gemseq/gemseqp/AA1995.DAT *
16:	/SIDS8/gcgcata/gemseq/gemseqp/AA1996.DAT *
17:	/SIDS8/gcgcata/gemseq/gemseqp/AA1997.DAT *
18:	/SIDS8/gcgcata/gemseq/gemseqp/AA1998.DAT *
19:	/SIDS8/gcgcata/gemseq/gemseqp/AA1999.DAT *
20:	/SIDS8/gcgcata/gemseq/gemseqp/AA2000.DAT *
21:	/SIDS8/gcgcata/gemseq/gemseqp/AA2001.DAT *
22:	/SIDS8/gcgcata/gemseq/gemseqp/AA2001.DAT *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1912	100.0	360	22	AA820492	Human cervical can
2	1071	56.0	203	22	AA626243	Human ribosomal pr
3	600	31.4	112	21	AA165193	Human 5' EST relat
4	337	17.6	162	21	AA858272	Lung cancer associ
5	124	6.5	586	16	AAW33891	Flea saliva protei
6	124	6.5	586	19	AAW82365	Flea saliva protei
7	94	4.9	428	20	AAW95086	S. pneumoniae resp
8	94	4.9	428	20	AAW95087	S. pneumoniae resp
9	92	4.8	662	18	AAW20939	H. pylori cytoplasm
10	89.5	4.7	230	22	AAW85022	Shrimp white spot
11	87	4.6	648	19	AAW75910	Helicobacter methi

12	86	4.5	2785	21	AAV57148	Human down-regulated
13	84.5	4.4	165	21	AAB842267	Human OREF-ORF2031
14	84	4.4	462	22	AAB94615	Human protein sequ
15	84	4.4	580	20	AAV35191	Chlamydia pneumoniae
16	84	4.4	1257	17	AAB87668	Alpha-ketoglutaric
17	84	4.4	1257	19	AAAM1781	B. lactofermentum
18	84	4.4	2150	21	AAV53898	Amino acid sequenc
19	84	4.4	2165	22	AAB90617	Human secreted pro
20	83.5	4.4	500	16	AAB84561	Human prostacyclin
21	82.5	4.3	603	20	AAV35593	Protein Involved I
22	82.5	4.3	778	21	AAAG51229	Arabidopsis thailia
23	82.5	4.3	806	21	AAAG51228	Arabidopsis thailia
24	82.5	4.3	927	21	AAAG51227	Arabidopsis thailia
25	81.5	4.3	237	16	AAB84562	Arabidopsis thailia
26	81.5	4.3	709	17	AAB99801	Human prostacyclin
27	80.5	4.2	242	21	AAB84253	CR11-7 nerve prote
28	80.5	4.2	344	21	AAAG24909	Human OREF-ORF2317
29	80.5	4.2	355	21	AAAG24908	Arabidopsis thailia
30	80.5	4.2	367	21	AAAG24907	Arabidopsis thailia
31	80.5	4.2	1545	19	AAAS3425	Arabidopsis thailia
32	80	4.2	200	15	AAAS3425	Rat sulphhytorea
33	80	4.2	200	20	AAAB3340	Human/rat chimera
34	80	4.2	503	22	AAV72161	Modified ciliary n
35	80	4.2	956	22	AAB79523	Human RNA metaboli
36	80	4.2	1257	22	AAAG30936	Corynebacterium glu
37	79.5	4.2	388	21	AAV67237	C glutamicum prote
38	79.5	4.2	557	21	AAAB53400	Barnacle fifth adhe
39	79.5	4.2	873	22	AAAB25699	Human colon cancer
40	79	4.1	452	21	AAB07182	Human protein sequ
41	79	4.1	753	21	AAV96159	Drosophila melanog
42	79	4.1	877	19	AAAB53828	Arabidopsis thailia
43	79	4.1	877	21	AAAB82601	Pseudomonas offiv
44	79	4.1	877	22	AAAB82258	Pseudomonas alcali
45	78.5	4.1	335	22	AAB94947	Pseudomonas alcali

ALIGNMENTS

RESULT 1

AAB20492
ID AAB20492 standard; Protein; 360 AA

XX AAB20492;
AC

XX	21-JUN-2001	(first entry)
DT		
TH		

Human cervical cancer 1 protooncogene-encoded protein.

AA Cervical cancer 1 protooncogene; HCCR-1; oncogene; human;
KW lung cancer; leukemia; lymphoma; kidney cancer; liver cancer;
KI lung cancer; leukemia; lymphoma; kidney cancer; liver cancer;

XX
KW
ovarian cancer; diagnosis; gene therapy.
XX
lung cancer; leukaemia, lymphoma, kidney cancer, liver cancer
XX

Homo sapiens.

	Key
FH	Peptide
FT	

Protein	/label = Signal_peptide
FT	26..360
FT	

FT	FT	Domain	/label= Mature_protein
FT	FT	Domain	26..142

FT	/label= Extracellular_domain
FT	143..162
FT	Domain

FT	/label= Transmembrane_domain
FT	163..360
Domain	

FT	/label= intracellular_domain
Modified-site	313..315

FI	Modified-site	42	note="Asn is N-glycosylated"
FT	Modified-site	42	
FM	Modified-site	42	note="O-mannosylated"

FT	/note=	0 phosphorylated
FT	Modified-site	48
FT		

FT	Modified-site	34	7/1000	0	phosphorylated

100

FT	Modified-site	/note= "N-myristylated"
FT	38	
FT	/note= "N-myristylated"	
XX	WO200127149-A1.	
XX		
XX	19-APR-2001.	
XX		
XX	30-MAR-2000; 2000WO-KR00284.	
XX		
XX	15-OCT-1999; 99KR-0044811.	
XX		
XX	(KIMJ/) KIM J W.	
XX		
XX	Kim JW;	
XX		
XX	WPI: 2001-290710/30.	
DR	N-PSDB: AAF30809.	
XX		
PT	Novel human cervical cancer 1 protooncogene is useful in the diagnosis	
PT	of various cancers, e.g., leukaemia, lymphoma, kinder, liver, lung,	
PT	ovary, and uterine cervix cancers -	
XX		
XX	Claim 3; Page 60-62; 67pp: English.	
XX		

CC The present sequence is that of the protein product of a novel
CC human cervical cancer 1 protooncogene, HCCR-1 (see AF038091).
CC The protooncogene was identified by differential display RT-PCR
CC analysis to isolate a clone that was expressed in cervical cancer,
CC metastatic tissue and CUMC-6 cervical cancer cells but not in
CC healthy tissue, and use of the isolated clone to screen a phase
CC lambda gII human lung embryonic fibroblast cDNA library. The
CC HCCR-1 protein is markedly hydrophobic and has a characteristic
CC single membrane-spanning domain and pre-secretory signal peptide.
CC The presence of a C-terminal glycosylation site suggests that
CC HCCR-1 is a type II membrane protein. HCCR-1 protooncogene is
CC overexpressed in cervical cancer tissues, cervical cancer cell
CC lines and lung cancer cell lines. Overexpression is also observed
CC in leukemia, lymphoma, kidney, liver and ovarian cancers. The
CC protooncogene is used in claimed kits for the diagnosis of cancer.
CC Antisense sequences are used in a claimed process for treating or
CC preventing cancer in humans. Also claimed are a vector comprising
CC the protooncogene, a microtubule transformed with the vector,
CC especially *Escherichia coli* JM109/HCCR-1 (KCMC 0661BP), and a
CC process for preparing HCCR-1 protein by culturing the host cells.

Query Match	100.0%;	Score 1912;	DB 22;	Length 360;
Best Local Similarity	100.0%;	Pred. No. 1.2e-200;		
Matches 360;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0

QY	1	MAISFCVMAKRSVWMSAVT	PGHEFTYRRLDLOGSGSLMGAPRSSKTLSPKADVUKMLSY	60
Db	1	MAISFCVMAKRSVWMSAVT	PGHEFTYRRLDLOGSGSLMGAPRSSKTLSPKADVUKMLSY	60
QY	61	VTKTKAINSGYHRLGRHP	PRFYLLVTYTEMKGLOMLMAAKKARIKTJPMKNKHTFQOL	120
Db	61	VTKTKAINSGYHRLGRHP	PRFYLLVTYTEMKGLOMLMAAKKARIKTJPMKNKHTFQOL	120
QY	121	PYKEMHLLKOPFODYTKCL	FTGIIISIPFPANNVYFLMLYFPROLLIRHEWPKOOPFL	180
Db	121	PYKEMHLLKOPFODYTKCL	FTGIIISIPFPANNVYFLMLYFPROLLIRHEWPKOOPFL	180
QY	181	DIYHAERKOSHEIITSYLEKY	IPRLSDAGLRMRLTDLCKRIORGNPAINDLALRECS	240
Db	181	DIYHAERKOSHEIITSYLEKY	IPRLSDAGLRMRLTDLCKRIORGNPAINDLALRECS	240
QY	241	NHPLGNNOIALHYKALS	RAMLLTSYLRPRLLRNHRTKTTPTVYHOLDALAKMLGIGOLTA	300
Db	241	NHPLGNNOIALHYKALS	RAMLLTSYLRPRLLRNHRTKTTPTVYHOLDALAKMLGIGOLTA	300

QY 301 QEVNSACYLKLGNSTHIGEDPRRTWLGWLQISCSLKKEELSLLHNVLSTNYIGTRR 360
 |||||
Db 301 qevnsacylqinsthigedrcrtwlgwlqiscslkeelsllhnvllstnyigrtr 360

RESULT	2
AA62443	
ID	AA62443 standard; Protein: 203 AA.
XX	
AC	AA62443;
XX	
DT	05-SEP-2001 (first entry)
XX	
DE	Human ribosomal protein L14.22.
XX	

KW	Human: ribosomal protein L14.22, malignant neoplasm; haemopathy;
KW	HIV infection; immunological disease; inflammatory disorder; cytostatic;
KW	haemostatic; virucide; immunomodulatory; antiinflammatory.
XX	
OS	Homo sapiens.

PN WO200138389-A1.

PD 31-MAY-2001.

PF 20-NOV-2000; 2000WO-CN00471.

PR 24-NOV-1999; 99CN-0124098.

PA (BIOR-) BIOROAD GENE DEV LTD SHANGHAI.

PI Mao Y, Xie Y,

DR WPI; 2001-355906/37
DR N-PEDD; 8AH45402

XX	Human ribosomal
DT	

PT diagnosis and treatment of malignant tumors, hemopathy, human immunodeficiency virus infection, immunological diseases and PT

pt Inflammation -
XX

PS Claim 1; Page 20-21; 30pp; Chinese
XX

cc This invention relates to human ribosomal protein L14.22 and cdna
cc encoding it. The invention includes a vector containing the cDNA, a host
cc encoding it. The invention includes a vector containing the cDNA, a host

cell transformed with the vector and an antibody directed against human ribosomal protein L14.22. The cDNA and protein sequences are used in the

immunodeficiency virus (HIV) infection, immunological diseases and

sequence may result in cytostatic; haemostatic; virucide;

CC human ribosomal protein L14.22

Sequence 203 AA;

Query Match	56.0%;	Score 1071;	DB 22;	Length 203;
Best Local Similarity	100.0%;	Pred. No. 6.3e-109;		
Matches 203; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0

QY 158 MYLFPROLLIRHFWTPKQOTDFLDIYHAEFKQSHPEIISYLEKVIPLISDAGLRWRLTDL 217

Db 1 mylfprqlllrhfwtpkqgtdfidlyhafkrkqshpelisylekviprlsdaglrwrldl 60

218 CTKIQRGTHPAIHDLALRECFNSNHP LGMNQLQALHVKALSRAMLLTSYLPPLLRHLK 277

Db 61 ctkigrthpaihdllalrecfsnhp1gmnglqalhvka1sraml1ttsy1pp11rhrik 120

278 THTTVIHQLDKALAKLGIGLTAEVKSACYLRLNSTHIGEDRCRTWLGEWLQISCSLK 337

Db 121 tttvindhakalalgigtagevksacylrglnsthigedrcrtwlgewlqiscsik 180

[illegible]

RESULT	5
AAW33891	
ID	AAW33891 standard; Protein; 586 AA
XY	

DT	14-APR-1998	(first entry)
XX		
DE		
XX		
KW	Flea saliva protein; fspw(F);	allergic dermatitis; therapy
KW	diagnosis; fspw(F)586.	
XX		
OS	Ctenocephalides felis.	

	Location/Qualifiers
FH	Key
FT	Misc-difference
FT	379
FT	/note= "encoded by AWG"

PN	WO9737676-A1.
XX	
PD	16-OCT-1997.
XX	
PF	10-APR-1997;
XX	
PR	10-APR-1996;
	96US-0630822
	97WO-US05959

PA (HESK-) HESKA CORP.
XX
PI Hunter SW, Slim G, Weber ER
XX
DR WPI, 1997-512409/47.
DR N-PSDB; AAT92839.

PT New flea saliva proteins - useful for treating allergic dermatitis
PT and as diagnostic reagents

PS Example 3; Page 127-129; 179pp; English.

This polypeptide comprises a non-full-length flea salivary protein, designated PfSPM(F)586, encoded by a cDNA clone (see 1928319), designated PfSPM(F)1758, isolated from a flea salivary gland library. PfSPM(F)586 has a calculated mol. wt. of about 66,547 and a calculated pI of 4.80. A Genbank homology search revealed no significant homology between PfSPM(F)1758 and PfSPM(F)586 and known nucleic acid or amino acid sequences. Novel nucleic acids (see 1928320-27) and novel flea salivary proteins (M30480-91), their fragments or mimotopes can be used in claimed methods for treating allergic dermatitis in an animal, for testing if an animal is susceptible to, or has, allergic dermatitis and for desensitising a host animal to allergic dermatitis.

SQ **Sequence** **586 AA;**

Query Match	6.5%;	Score 124;	DB 18;	Length 586;
Best Local Similarity	27.0%;	Pred. No. 0.00014;		
Matches 37;	Conservative 29;	Mismatches 69;	Indels 2;	Gaps 2

```

QY 219 TKI-ORGHPIIHIDILALRECFSNHPIRGANOQALAHVAKALSSAMLETSTYLRPELIRHRK 277
      ||: ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: : ||: :
Db 34 tkmrysgsgaqaneimfkfede-lltdslarplvalcvleislgtntfifqrl 92
QY • 278 THHTVTHQDLKALAKLIGOLTQAEVKSACVLRGLNSTHIEGDRRTWLGEMQDISCLK 337

```

Db 93 mkrlrslaaddkmkgivsmlysevggaactargmazympehrlrrgledwlnslnek 152

Oy 338 EAELSLLLNHVLLSTN 354

Db 153 vppslilllsralmpen 169

RESULT	6
AAW82365	
ID	AAW82365 standard; Protein; 586 AA
XX	
AC	AAW82365;
XX	

26-APR-1999 (first entry)
Flea saliva protein Pfspm(F)-586

Flea saliva protein: PfSPM(F)-586; allergic dermatitis; allergen allergy; therapy; diagnosis; vaccine; ectoparasite.

Ctenocephalides sp.

Key	Location/Qualifiers
Misc-difference	379
	/note= "encoded by AWG"

W09845408-A2.

15-OCT-1998.

15-OCT-1997; 97WO-US18669

10-APR-1997; 97WO-US05959.

(HESK-) HESKA CORP.

Weber ER;

WPI; 1998-594480/50.
N-PSDB; AAV73387.

New isolated ectoparasite saliva genes - used to develop products for the diagnosis, prevention, treatment and determining susceptibility to allergic dermatitis

Example 3; Page 99-101; 172pp; English.

This non full-length polypeptide comprises the 66 kDa flea saliva protein nFspm(F)-556, as deduced from cDNA clone nFspm(F)-1758 (see AAV73387) . It shows no significant homology to database sequences. The invention is directed to novel products and methods for isolating ectoparasite saliva proteins (ESPs) . It provides ESps (see AAM2382-93), nucleic acid molecules encoding them, methods for their recombinant production, and therapeutic compositions for treating allergic dermatitis that comprise at least one ESP, as well as assay kits for testing if an animal has, or is susceptible to, allergic dermatitis, and a method of desensitising a host animal to allergic dermatitis using ESps. The ESps can also be used for the production of antibodies useful in diagnosis or in vaccines for passive immunisation against allergic dermatitis. The ESP may be flea saliva protein and the allergic dermatitis may be flea allergy dermatitis.

50 Sequence 586 AA;

Query Match	6.5%;	Score 124;	DB 19;	Length 586;
Best Local Similarity	27.0%;	Pred. No. 0.00014;		
Matches 37;	Conservative 29;	Mismatches 69;	Indels 2;	Gaps 2

219 TKI-ORGTHPAIDILALRECFSNPLGMNQLALHVKALSRAMULTSYLPPPLRHRK 277

Db 34 tkmrsgagaganeimkfsklfedeltdlsarpqvalorcleistgltnflrfqlr 92
 QY 278 THTTVIHOLDKALAKIGTOLTAQAEYKSAQYRGLNTHIGEDRCRTWLGWMLQISCLK 337
 Db 93 mklrtaaddkmqkgvlsmtcyevgacrararntaygmpehrlrrqledwlnslnek 152
 QY 338 EAEISLLHNHVLSTN 354
 Db 153 vppslllslraalnpen 169

RESULT 7
 AAM95086
 ID AAM95086 standard; Protein; 428 AA.
 AC AAM95086;
 DT 20-MAY-1999 (first entry)
 DE S. pneumoniae response regulator polypeptide.
 KW Response regulator; two component signal transduction system; TCSTS;
 KW antibacterial; genetic immunisation; gene therapy; bacterial adhesion;
 KW wound; body implant; bacterial infection; otitis media; conjunctivitis;
 KW pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis;
 KW meningitis.
 OS Streptococcus pneumoniae.
 PN EP900846-A2.
 PD 10-MAR-1999.
 PF 02-SEP-1998; 98EP-0307054.
 PR 09-SEP-1997; 97US-0060714.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Biswas S, Ge Y, Holmes D, Ingraham K, Throup J;
 PI Wallis N, Zalacain M;
 DR WPI: 1999-155939/14.
 DR N-PSDB: AAX26104.
 XX New Streptococcus pneumoniae Response Regulator (RR) polypeptide and
 PT polynucleotide - useful as diagnostic reagents and for prevention
 PT and treatment of Streptococcus infections which cause
 PT conjunctivitis, sinusitis and meningitis
 XX Claim 4: Page 5; 56pp: English.
 XX This represents a Streptococcus pneumoniae response regulator protein,
 CC which is part of the two component signal transduction system (TCSTS).
 CC Host cells containing a vector comprising the nucleic acid are used for
 CC the recombinant expression of the protein. The response regulator
 CC sequences are useful for diagnosing a disease or susceptibility to a
 CC disease related to the polypeptide levels. They can diagnose the stage
 CC and type of infection. They are also useful for screening for compounds
 CC which stimulate or inhibit polypeptide function. Agonists and antagonists
 CC are useful for treatment of conditions associated with response regulator
 CC imbalance, and are therefore potential antibacterial compounds. The
 CC polynucleotides are useful for genetic immunisation and in gene therapy
 CC and antisense CC sequences are useful for inhibition of expression of
 CC the DNA. The response regulator polypeptides can prevent adhesion of
 CC bacteria to matrix proteins, and are useful for use on wounds and body
 CC implants to prevent bacterial infection. Anti-response regulator
 CC antibodies induced by the polypeptide are useful for preventing or
 CC treating infections, especially bacterial infections, and also for
 CC isolating clones expressing the polypeptide. Diseases prevented,
 CC diagnosed and treated include those caused by bacterial infection,
 CC especially Streptococcus pneumoniae infections, which cause otitis
 CC conjunctivitis, pneumonia, bacteremia, sinusitis, pleural empyema,

CC endocarditis and especially meningitis.
 XX Sequence 428 AA:
 SQ
 Query Match 4.9%; Score 94; DB 20; Length 428;
 Best local Similarity 20.8%; Pred. No. 0.17;
 Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;
 QY 29 QLCRSGIAMGAPRSSKLIHSFKADYKNLMSYVVTKTALNKGTHRFLGRHPPREYLYTI 88
 Db 122 qlgerg-----kqsqtlsgeldeagrtvsgkenwigskeqg-----stflpyyv 170
 QY 89 FMKGLOLMADAKKARRIKTNMKWNHNIKFHOLPYREMEHLRQFODVTKCLEFLGIIISPP 148
 Db 171 lqdgwqflfisdqpldglvrvpf-----eayqg-ehferwklnaektlfysvsnq 220
 QY 149 ---FA-----NTLVPLMLLPFR-----QLLR-----HPWT 172
 Db 221 seslfayeplyrvlqgnlnqiveelnllekvlenrcpripkqifqfmdvfnlfe 280
 QY 173 PKQOTDFLDI---YHAFRKOSHPEIISYLEKVIPLISDAGLRWRLDLCIKIQRGTHPAI 229
 Db 281 hlkdadmtdivktlhal--gsfdelvyiket--liffg-gyrme----- 322
 QY 230 HDITLRCFESNHPGLGNNOLOALHVKALSRAMLITSTLPPLRHLKTHHTYIHOLDKA 289
 Db 323 -nvsvlev-----lgrdygkelskalskalflnpyvjqlklr--eclstlaelnkq 374
 QY 290 LAKLIGTQL--TAQEVKSACVLRGLNS 314
 Db 375 rikaagqlllstdslenicyavgsn 401

RESULT 8
 AAM95087
 ID AAM95087 standard; Protein; 428 AA.
 AC AAM95087;
 DT 20-MAY-1999 (first entry)
 DE S. pneumoniae response regulator ORF sequence.
 KW Response regulator; two component signal transduction system; TCSTS;
 KW antibacterial; genetic immunisation; gene therapy; bacterial adhesion;
 KW wound; body implant; bacterial infection; otitis media; conjunctivitis;
 KW pneumonia; bacteremia; sinusitis; pleural empyema; endocarditis;
 KW meningitis; ORF; open reading frame.
 OS Streptococcus pneumoniae.
 PN EP900846-A2.
 PD 10-MAR-1999.
 PF 02-SEP-1998; 98EP-0307054.
 PR 09-SEP-1997; 97US-0060714.
 PA (SMIK) SMITHKLINE BEECHAM CORP.
 PI Biswas S, Ge Y, Holmes D, Ingraham K, Throup J;
 PI Wallis N, Zalacain M;
 DR WPI: 1999-155939/14.
 DR N-PSDB: AAX26105.
 XX New Streptococcus pneumoniae Response Regulator (RR) polypeptide and
 PT polynucleotide - useful as diagnostic reagents and for prevention
 PT and treatment of Streptococcus infections which cause
 PT conjunctivitis, sinusitis and meningitis
 XX

PS Claim 19; Page6; 56pp; English.

CC This represents a Streptococcus pneumoniae response regulator protein,
CC which is part of the two component signal transduction system (TCSTS).
CC Host cells containing a vector comprising the nucleic acid are used for
CC the recombinant expression of the protein. The response regulator
CC sequences are useful for diagnosing a disease or susceptibility to a
CC disease related to the polypeptide levels. They can diagnose the stage
CC and type of infection. They are also useful for screening for compounds
CC which stimulate or inhibit polypeptide function. Agonists and antagonists
CC are useful for treatment of conditions associated with response regulator
CC imbalance, and are therefore potential antibacterial compounds. The
CC polynucleotides are useful for genetic immunisation and in gene therapy
CC and antisense CC sequences are useful for inhibition of expression of
CC the DNA. The response regulator polypeptides can prevent adhesion of
CC bacteria to matrix proteins, and are useful for use on wounds and body
CC implants to prevent bacterial infection. Anti-response regulator
CC antibodies induced by the polypeptide are useful for preventing or
CC treating infections, especially bacterial infections, and also for
CC isolating clones expressing the polypeptide. Diseases prevented,
CC especially Streptococcus pneumoniae infections, which cause otitis media,
CC conjunctivitis, pneumonia, bacteremia, sinusitis, pleural empyema,
CC endocarditis and especially meningitis.

CC
XX
SQ Sequence 428 AA:

Query Match 4.9%; Score 94; DB 20; Length 428;
Best Local Similarity 20.8%; Pred. No. 0.11;

Matches 68; Conservative 56; Mismatches 115; Indels 88; Gaps 16;

```

QY 29 QLRGSLAMGAPRSSKLLSPKADVKNLMSYVVTKAINGKYRFLGRHPRFYLYTI 88
DB 122 qlgery-----ksgqlsgeldegfvsylgdkenwvlglskekq-----stflpyyv 170
QY 89 FMKGLQMLMADAKKARRIKTMNMKNHNFHQLPYREMEHLROFRODYTKCLFLGITSIP 148
DB 171 lggawqifisdqldjlvtrp-----eapryg--ehetwkylnckllfyssvnlqg 220
QY 149 -----FA-----NYLVFLMYLPPR-----QLLR-----HFWT 172
DB 221 seelfayeyeplyvllglnlqveelnllekvlenltpirltkqlfqlgmndvfhle 280
QY 173 PKQOTDFLDT---YHAFRKOSHPEIISYLEKVIPLISDGLRMLDLCIKIQRGTHNPAI 229
DB 281 hlkadmdldvktihai--gsfdeivsyiket--liffg-qyrme-----322
QY 230 HDILALRECFENHGLGMNOLOALHVKALSRAMLITSLPRPLIRNLKTHHTVVIHOLDKA 289
DB 323 -nvsvslev-----lgrdygkelskdkalkalfinpvyigqlikr--elcstfaellnkq 374
QY 290 LAKLIGQL--TAQEVKSACYLRGLNS 314
DB 375 rikaaqqlilstdslenlcayavgsn 401

```

RESULT 9

AAW20999
ID AAW20999 standard; protein; 662 AA.

AC AAW20999;

DT 21-JUL-1997 (first entry)

DE H. pylori cytoplasmic protein, hp4el4522orf11.

KW Cytoplasmic; vaccine; prevention; treatment; infection; identification;
KW binding compound; bacterium; life cycle; activator; bacteria; inhibitor;
KW duodenal ulcer disease; chronic gastritis; diagnosis; envelope.

XX Helicobacter pylori.
OS
XX

PN WO6640893-A1.

XX 19-DEC-1996.

XX 06-JUN-1996; 96MO-US09122.

XX 01-APR-1996; 9605-0630405.

XX 07-JUN-1995; 9505-0487032.

XX (ASTR) ASTRA AB.

PI Berglinth OT, Smith D, Mellgaard BJ.

DR WPI; 1997-052306/05.

XX N-PDB; AAT68252.

PT Helicobacter pylori nucleic acid sequences and related
PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
PT infection, and to detect Helicobacter

PS Claim 61; Page 1389-1390; 1481pp; English.

CC The present sequence is a H. pylori cytoplasmic protein involved in
CC mRNA translation and ribosome biogenesis.
CC The protein may be used in a vaccine to prevent or treat H. pylori
CC infection or to identify H. pylori polypeptide binding compounds.
CC The genomic sequence of H. pylori life cycle activators or inhibitors.
CC DNA. The sequences were analysed by mechanically shearing the bacterial
CC and the predicted coding regions defined by computer evaluation. To
CC identify likely H. pylori antigens for vaccine development, the amino
CC acid sequences predicted from various ORF were analysed for significant
CC homology to other known or exported membrane proteins. Having identified
CC and determined the sequences of interest, particular regions can be
CC isolated from H. pylori by PCR amplification for recombinant polypeptide
CC production, e.g. in E. coli hosts.

XX
SQ Sequence 662 AA:

Query Match 4.8%; Score 92; DB 18; Length 662;
Best Local Similarity 17.7%; Pred. No. 0.53;

Matches 49; Conservative 51; Mismatches 113; Indels 64; Gaps 8;

```

QY 28 IOLRSGIAMGAPRSSKLLSPKADVKNLMSYVVTKAINGKYRFLGRHPRFYLYTI 87
DB 210 lsirtsfewgdlpdkmm-dpkhvvyywldallnyasalq-----249
QY 88 FMKGLQMLMADAKKARRIKTMNMKNHNFHQLPYREMEHLROFRODYTKCLFLGITSIP 147
DB 250 -ynglnlknahferarhi--vgkdllrfhalywp-----afinsindp 290
QY 148 PFANYLVFLMYLPPRQLIRHFWT-----PKQOTDFLDTIYHAFRKOSHPEIISYLEKV 201
DB 291 lf-----kqlcvhgwltlegvkmkslgnvldagklameyieeltryfllire 337
QY 202 IPLISAGLRMR--LNDLCTKIQRGTHNPAIHDLALRECFENHGLGMNOLOALHVKALSR 259
DB 338 vpfqgdgdfekkalverinanlndlglnlrlgmakkyfynkskclacaypkelek 397
QY 260 AMLL-----TSYLPRPLIRNLKTHHTVVIHOLDKALAK 292
DB 398 ahqldmansfvpkmqjhlkaleelfnydyflnklak 434

```

RESULT 10

AAG85022

ID AAG85022 standard; Protein; 290 AA.

AC AAG85022;

DT 11-SEP-2001 (first entry)

XX Shrimp white spot Bacilliform virus (WSBV) protein 113.
 DE Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
 XX antiviral agent; gene expression; antisense construct;
 KW transgenic viral resistant shrimp.
 KW white spot syndrome virus.
 OS WO200138351-A2.
 PN 31-MAY-2001.
 XX 08-NOV-2000; 2000WO-US28888.
 PD 24-NOV-1999; 99CN-0124717.
 XX (PENY-) PE CORP NY.
 PA (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANIC C. A.
 PA (SINO-) SINOGENOMAX CO LTD.
 XX Xu X, Yang F, He J, Pham L, He M, Ye Y, Shen Y, Kodira C;
 PI WPI: 2001-355877/37.
 DR N-PSDB: AAH62802.
 XX Primary nucleotide sequence of the shrimp white spot Bacilliform virus
 PT (WSBV), useful for producing viral polypeptides that can be used to
 PT screen for agents that are useful for treating WSBV infection -
 PS Claim 1; Figure 3; 626pp; English.
 XX The invention provides the primary nucleotide sequence of the WSBV genome
 CC (AAH62889), predicted transcript sequences (AAH62689-AAH62839) and
 CC encoded proteins (AAH62891-AAH62901) and oligonucleotide sequences
 CC (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
 CC molecules and proteins of the invention are useful for diagnosis and
 CC monitoring viral infection. In screens for antiviral agents and for
 CC monitoring viral gene expression or activity during a treatment regimen.
 CC The nucleic acid molecules are also useful as antisense constructs to
 CC control viral gene expression in infected cells and tissues and to create
 CC transgenic viral resistant shrimp.
 CC Sequence 290 AA:
 SQ
 Query Match 4.7%; Score 89.5; DB 22; Length 290;
 Best Local Similarity 26.9%; Pred. No. 0.3;
 Matches 36; Conservative 21; Mismatches 42; Indels 35; Gaps 5;
 QY 48 SPRADYKNTMSYVYTKAINGKRYHFLGRHPRFYI-----LYTIPMK 91
 Db 122 sqgkdvgeyvkikvkektlml-----frgfyspyilkslpdipfgekerfvstdlil 175
 QY 92 GLQMLWADA-KKARIRIKTNMKNHINIKFQULPYREMEHLROFRODYTKCLFLGIISIPPA 150
 Db 176 glgfsadwemkilaiegmrksklyltwvpaevchlkkygd-----lvnplfk 227
 QY 151 NY---LVPLMLYL 160
 Db 228 synhcivldlyl 241
 RESULT 11
 AAW75910 ID AAW75910 standard; Protein; 648 AA.
 XX AAW75910;
 XX 12-NOV-1998 (first entry)
 DT Helicobacter methionyl-tRNA synthetase.
 XX

KW Helicobacter; aminocyl-tRNA synthetase; isoleucyl-tRNA synthetase;
 KW methionyl-tRNA synthetase; leucyl-tRNA synthetase; valyl-tRNA synthetase;
 KW lysyl-tRNA synthetase; seryl-tRNA synthetase; human gastric mucosa;
 KW chronic gastritis; antimicrobial compound; antisense inhibition;
 KW H. pylori; bacterium; enzyme.
 OS Helicobacter pylori.
 XX US5801013-A.
 PN 01-SEP-1998.
 XX 26-MAY-1995; 95US-0451715.
 PR 26-MAY-1995; 95US-0451715.
 XX (CUBI-) CUBIST PHARM INC.
 XX Houman F, Qiu Y, Schimmel PR, Shen X, Tao J;
 PI WPI: 1998-494769/42.
 DR N-PSDB: AAV53135.
 XX DNA encoding Helicobacter amino-acyl-tRNA synthetase proteins
 PT useful for recombinant production of the enzyme, anti-sense
 PT constructs and hybridisation probes, and construction of tester
 PT strains to test for inhibitors of the enzyme
 PS Claim 7; Columns 61-66; 76pp; English.
 XX This represents a Helicobacter methionyl-tRNA synthetase. The invention
 CC provides nucleic acid sequences (AAV53134 to AAV53139) encoding
 CC Helicobacter aminocyl-tRNA synthetases (hATTRNs) (AAW5909 to AAW5914),
 CC where the aminocyl-tRNA synthetase is isoleucyl-tRNA synthetase,
 CC methionyl-tRNA synthetase, leucyl-tRNA synthetase respectively and at least
 CC a portion of the hATTRNs has catalytic activity or binding function. The
 CC nucleic acids and recombinant cells containing the nucleic acids can
 CC also be used as hybridisation probes in experiments to identify hATTRNs
 CC sequences. H. pylori is bacteria that infects human gastric mucosa,
 CC leading to chronic gastritis. The amino acyl tRNA synthetases are used by
 CC the bacterium during protein synthesis, and as such can be a target for
 CC antimicrobial compounds. The tester strains can be used to screen for
 CC such compounds, and antisense constructs based on the nucleic acids can
 CC be used for antisense inhibition of the enzyme. The proteins themselves
 CC can be used to raise antibodies, which in turn can be used for
 CC purification and study of the enzyme. Screening of inhibitors of the
 CC enzyme, as well as antisense antimicrobial activity through the nucleic
 CC acids are novel approaches to inhibition activity against the bacteria,
 CC as many present antibiotics used have side-effects and the bacteria are
 CC becoming resistant to them.
 CC Sequence 648 AA:
 SQ
 Query Match 4.6%; Score 87; DB 19; Length 648;
 Best Local Similarity 18.1%; Pred. No. 1.8;
 Matches 50; Conservative 49; Mismatches 114; Indels 64; Gaps 8;
 QY 28 LQGRSGIAGAPRSSKRLSPRADYKNTMSYVYTKAINGKRYHFLGRHPRFYI 7
 Db 204 lsirtsfewgiprpkmn-dpkhvyvwdallnyasalg-----afmslmlp 243
 QY 88 IFMKGLQMLWADAKKARIRIKTNMKNHINIKFQULPYREMEHLROFRODYTKCLFLGIISIP 147
 Db 244 -ylngldnkmahfecahti---vgkdllrthaalyp-----afmslmlp 284
 QY 148 PFANYLVFLMYLFRPOLIRHEWT-----PQQQDFLDIYHAFKQSHPELIISLEKV 201
 Db 265 lf-----kqlcvhgwvlllegvkmksklsignvdagklameygleelyfllire 331
 QY 202 IPLISDAGLRMR--LTDLCTKIQNGTTPAIDHIALRECSNHPLOMNOQLAHVAKLSR 259

Db 332 vfgdgdfskkajierinanindgnllnrlgmakkyfnylsakltaysskelek 391
 QY 260 AMLL-----TSLPPLRLRHLKHTTYIHOLDKALAK 292
 Db 392 ahqildansstvpkmqlhkaeeefnyvdfinkxiak 428

RESULT 12
 ID AAY57148 standard; protein; 2785 AA.
 AC AAY57148;

DE 28-FEB-2000 (first entry)
 DE Human down-regulated in metastasis (DRIM) amino acid sequence.
 KW Down-regulated in metastasis; DRIM; human; antimetastatic activity;
 KW antibody; tumour; treatment; therapy.

OS Homo sapiens.

PN WO960116-A1.

PD 25-NOV-1999.

PF 17-MAY-1999; 99WO-EP03396.

PR 18-MAY-1998; 98EP-0303895.

PA (HOPE) ROCHE DIAGNOSTICS GMBH.

PI (ISIS-) ISIS INNOVATION LTD.

PI Weidle U, Tarin D;

DR WPI: 2000-053296/04.

DR N-PSDB; AA245136.

PT New polypeptide with antimetastatic activity, useful for therapeutic
 PT compositions for tumour therapy

PS Claim 2; Page 42-49; 54pp; English.

CC This is the amino acid sequence of the human down-regulated in metastasis
 CC (DRIM) protein. The protein has antimetastatic activity. The DRIM protein
 CC is active in both its glycosylated and unglycosylated form, and can be
 CC produced by recombinant technology in prokaryotic cells. DRIM mRNA is
 CC strongly expressed in heart, skeletal muscle, pancreas, testis and ovary
 CC tissues. The nucleotide and protein sequences can be used to create
 CC anti-DRIM antibodies. The nucleic acids are useful in therapeutic
 CC compositions, especially for treating tumours. They are also useful for
 CC activating polynucleotides from the 5' untranslated region in gene
 CC therapy.

XX Sequence 2785 AA;

Query Match 4.5%; Score 86; DB 21; Length 2785;
 Best Local Similarity 19.8%; Pred. No. 19;
 Matches 62; Conservative 49; Mismatches 94; Indels 108; Gaps 14;

QY 42 SSKHLSPKADYKMLSYVVTTKALNGKYHPLGRHPR-FYLLTTPMKGLQMLMADA 100
 Db 901 ssqkkktrraaakqlahlvfksfn-----pralyleeklyelqlllmgd 949

QY 101 KKAIRIKTN---MKHNIKPHLPYRE-MEHL---RQFRDVTKCLF---GISIPPPA 150
 Db 950 qmqqkltldcmlyknp---hvpryrengrlledstfkeelvhfsiscnawnvxtahra 1006

QY 151 NYVFLMLYF-----PROLLIRHFWT---PKQQLDFLD-----LYH 184
 Db 1007 dlfpilmrillygmknktgsktgkaasgtlmaivlrflagrqpeidqfildllefpyrh 1066

QY 185 AFRKSHPELISYLE-----KVPLISDAGLRKRLTDLCTKRGTHPAIHDLALRECF 239
 Db 1067 fkgcecsavqavedldskvplgrqhg1----- 1097

QY 240 SNHPLGNMLOALHVKALSRAMLTSLYPLRLRHLKHTTYIHOLD----- 287
 Db 1098 -----lnslelvknshlsay1-pkllqlllcmatavshldgkexlqlrfing 1147

QY 288 -KALAKIGTOLT 299
 Db 1148 lknllrlgikmvt 1160

RESULT 13
 ID AAB42267 standard; protein; 165 AA.
 AC AAB42267;

DE 08-FEB-2001 (first entry)

DE Human OREFX OREF2031 polypeptide sequence SEQ ID NO:4062.

KW Human; open reading frame; OREFX; detection; cytostatic; hepatotropic;
 KW vulnery; antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotrophic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antineoplastic disease; coagulation;
 KW thrombosis; contraceptive.

OS Homo sapiens.

PN WO200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shimkets RA, Leach M;

DR WPI: 2000-602362/57.

DR N-PSDB; AAC76476.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -

PS Claim 11; Page 3250-3251; 5507pp; English.

CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human OREFX open reading frames 1 to 3161. The OREFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiproliferative; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotrophic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antineoplastic;

CC antihypertensive; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

CC Sequence 165 AA:

Query Match 4.4%; Score 84.5; DB 21; Length 165;
 Best Local Similarity 26.6%; Pred. No. 0.47;

Matches 41; Conservative 28; Mismatches 60; Indels 25; Gaps 8;

QY 214 LNDLCIKRGTHPAHDLALRECFSNHPLGMNOLA---LHVVALSRAMLLTSTL--- 267

DB 1 mthlnsyngdthpaindy--lwacalnshlgknealipivksvkcavtsdltrc 58

QY 268 ---PPLLRHRLKTHVTYHOLDKALAKLGOLTAQEVKSACYLRLGNS--THIGEDRC 322

DB 59 tlttgmwylhtrmsgkmsldka---prrqll--datigaylnthsrthsprry 112

QY 323 RTWIGEMLIQISCSLKEAELSLHLHNVLLSTNYL 356

DB 113 sefl-efts----karetfmahdghlqftqfi 140

RESULT 14

AAB94615 AAB94615 standard; Protein; 462 AA.

AC AAB94615;

DT 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:15462.

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.

OS Homo sapiens.

PN EP1074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-0116126.

PR 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

PA (HELI-) HELIX RES INST.

PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

DR WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602

PT full-length cDNAs defined in the specification, and for the detection

PT and/or diagnosis of the abnormality of the proteins encoded by the

PT full-length cDNAs -

PS Claim 8; SEQ ID 15462; 2537bp + CD ROM; English.

XX The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

CC Sequence 462 AA:

Query Match 4.4%; Score 84; DB 22; Length 462;
 Best Local Similarity 19.7%; Pred. No. 2.4;

Matches 60; Conservative 48; Mismatches 87; Indels 110; Gaps 18;

QY 87 TTPKGLQMLVAD--AKKARRIKTNMKNHNIKFPQLPYREMEHLRQPRODY---TKCLF 140

DB 4 tvekefimmctdyvstfnkimmwkgcgynpnyfhei--lkleeeeksasctky 61

QY 141 LGIISIPFAN-YVFLMLYLPFROLIRHWPTRKQDTPFDIYHARKOSHPEIISYLE 199

DB 62 -----ftntfildlvcelfgr-----arenfkemhrakfrandp--vnyle 101

QY 200 -----KVLPLISDAGLRWRLTDLCTKI---QRGTHP 227

DB 102 skkdffsfkiscgatslktfdllywkltpavst--lwe--dntfklaqdmratcp 157

QY 228 ATRD-----ILALRECFSNHPLGMNOLAHLVKALSRAMLLTSTYLPPLLRH-- 274

DB 158 afngnrklehllfslaeenfdny-----weyln-----nsksffrylknhikykcs 207

QY 275 -----RLKTHVTYHOLDKALAKLGOLTAQEVKSACYLRLGNS--THIGEDRCRTWLGW 329

DB 208 dngekmktf-----feksl-----dkntllsahestsvakdkst--asew 250

QY 330 LQISC 334

DB 251 ldlfc 255

RESULT 15

AAV35191 AAV35191 standard; Protein; 580 AA.

AC AAV35191;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae lipoprotein sequence.

KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;

KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;

KW vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN W09927105-A2.

XX 03-JUN-1999.
 PD
 XX
 PF 20-NOV-1998; 98MO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffais R;
 XX
 DR WPI; 1999-357842/30.
 XX
 PT
 XX
 PS Genome sequence of Chlamydia pneumoniae
 XX
 CC Page 1049-1050; Disclosure: 1912pp; English.
 CC
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAY31990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 CC
 XX
 SQ Sequence 580 AA;

Query Match 4.4%; Score 84; DB 20; Length 580;
 Best Local Similarity 23.7%; Pred. No. 3.3;
 Matches 44; Conservative 23; Mismatches 59; Indels 60; Gaps 8;
 QY 178 DFLDIYHAFKQSH-----PEIISYLEK-----VIPLI--SDAGLR 211
 Db 261 dldivekrgesfynpflpeiledlekglitvsndakcvfheafsfmvgksdgyn 320
 QY 212 WRLDLCIKIQRGTHPAIHDLALRECESNPTGSMNOLOALHVKALSRAMILTSYLPPL 271
 Db 321 yatdlaamyriedhadtllivtd-----lg---qslhtgldeataaagylqp91 370
 QY 272 LRH-----RLKTHP-----TVIHOLDKALAKLGIGOLTAEVKSACY 308
 Db 371 fshvgfglvdpggkklktrsgenvklrelltdaikaeealre-hrpeltdaigerap 429
 QY 309 LRGINS 314
 Db 430 vlgina 435

Search completed: December 20, 2001, 10:11:37
 Job time: 125 sec

Qy	1	MLSLRVCWARSXVWMSAVT	PCGHFTVR	RLOLGRSLG	LANCAPSS	LTHLS	SPRADYK	NLMSTY	60
Db	1	MLSLRVCWARSXVWMSAVT	PCGHFTVR	RLOLGRSLG	LANCAPSS	LTHLS	SPRADYK	NLMSTY	60
Qy	61	VAKTKA	INAKYARFGRHPRF	LYLTJEMFKGLQ	LYMDAKKARIRIKTNMKNH	IKFHOL			120
Db	61	VAKTKA	INAKYARFGRHPRF	LYLTJEMFKGLQ	LYMDAKKARIRIKTNMKNH	IKFHOL			120
Qy	121	PYREMHNLROFQDUTYK	CLFLG	ISIPPRANTV	LYMLTLP	RROLLRHNEFTPRQOOF	DFL		180
Db	121	PYREMHNLROFQDUTYK	CLFLG	ISIPPRANTV	LYMLTLP	RROLLRHNEFTPRQOOF	DFL		180
Qy	181	DIYHARFKOSHEPILIS	ULEKVIPLISDAG	LWRLTDL	CTQIGSTH	PAIDILAL	RECFS		240
Db	181	DIYHARFKOSHEPILIS	ULEKVIPLISDAG	LWRLTDL	CTQIGSTH	PAIDILAL	RECFS		240
Qy	241	NPELGNQLOALHVKAL	YSAMMLT	STIPRRLR	NRHNLKHTTV	IYHQDLAK	LGLGOLTA		300
Db	241	NPELGNQLOALHVKAL	YSAMMLT	STIPRRLR	NRHNLKHTTV	IYHQDLAK	LGLGOLTA		300

Db 241 NHPGMMQOLAHVKALSRAMLTSYLPRLHRLKHTHTVTHQDLKALAKLGIGQLTA 300
 QY 301 QEVKSACTGLGNTSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLGTTR 360
 Db 301 QEVKSACTGLGNTSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLGTTR 360

RESULT 2

QY3X3 PRELIMINARY: PRT: 100 AA.

ID 09Y3X3
 AC 09Y3X3
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, last sequence update)
 DE HYPOTHETICAL 11.3 KDA PROTEIN (DKFZP586A011.PROTEIN).
 GN DKFZP586A011.
 OS Homo sapiens (Human)
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=UTERUS;
 RA Koehler K., Beyer A., Mewes H.W., Gassenhuber J., Wiemann S.,
 RL Submitted (Mar-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=LUNG CARCINOMA;
 RA Strausberg R.,
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL050286; CAB43387.1;
 DR EMBL: BC000395; AAH00395.1;
 KW Hypothetical protein.
 SQ SEQUENCE 100 AA: 11318 MW: 4C5179D6F75156C7 CRC64;

Query Match 27.2%; Score 520; DB 4; Length 100;
 Best Local Similarity 100.0%; Pred. No. 6e-41;
 Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 MLTSTYLPRLHRLKHTHTVTHQDLKALAKLGIGQLTAQEVKSACTGLGNTSTHIGED 320
 Db 1 MLTSTYLPRLHRLKHTHTVTHQDLKALAKLGIGQLTAQEVKSACTGLGNTSTHIGED 60
 QY 321 RCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLGTTR 360
 Db 61 RCRTWLGEMLOISCSLKEAEELSLHNVLLSTNYLGTTR 100

RESULT 3
 P91617 PRELIMINARY: PRT: 436 AA.

ID P91617
 AC P91617
 DT 01-MAY-1997 (TREMblrel. 03, Created)
 DT 01-MAY-1997 (TREMblrel. 03, last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
 DE ANON-66DB PROTEIN.
 GN CG5989 OR ANON-66DB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SHAHINAV (WILD TYPE); TISSUE=WHOLE FLY;
 RA MEDLINE=9737971; Pubmed=9236770;
 RA Seebode-Larsen S., Urbaneczyk Moheli B., Lambertson A.;
 RT "The Drosophila ribosomal protein L14-encoding gene, identified by a
 RT novel minute mutation in a dense cluster of previously undescribed
 RT genes in cytogenetic region 6D.";
 RL Mol. Gen. Genet. 255:141-151(1997).

DR EMBL: Y10015; CA71122.1;
 DR FlyBase: FBgn0017429; CG5989.
 SQ SEQUENCE 436 AA: 50299 MW: 19A5EC0B5D71A5C CRC64;

Query Match 17.9%; Score 342; DB 5; Length 436;
 Best Local Similarity 26.6%; Pred. No. 1.3e-23;
 Matches 81; Conservative 67; Mismatches 148; Indels 8; Gaps 4;

QY 55 NLMSYVTKTKKAINGVHFFLGHPFRFYLTITNKGLOIMADAKKARIK--TNMK 112
 Db 125 NMODYIFTRFNVVKNVDVLEKNPKAMQLYRVEDGKDFEGDMKRPFLARIANDSP 184
 QY 113 HNIKFPOLPREMHNLPORODVTKCLFGLISIPFANVYLFVLMYFPQOLLIRHWT 172
 Db 185 QGIR--ALNKGLELYVMQPRDMKVAPALLICSLPVMGYAFPLVETYPSPFLTAHWT 242
 QY 173 PKQGTPELDIYNAFKQSPHELIISTLEKVIPLISAGLRMRDLCTKIQGTGPAIHDI 232
 Db 243 PQORSEFQSYTKRKRCNNKDVFCLOAKKATASHPKNSAFADILGQSGTPTPEML 302
 QY 233 LALRECFNHPGLMNQOLAHVKALSRAMLTSYLPRLHRLKHTHTVTHQDLKALAK 292
 Db 303 IDVKDIFAGCPYSLGMSKRVNVLNHLGPSI--FKHRLHEHAFVHYMDQATIR 359
 QY 293 L-CIGQLTAQEVKSACTGLGNTSTHIGEDRCRTWLGEMLOISCSLKEAEELSLHNVLL 351
 Db 360 EGVVHNLTPDALRYSCYLRGLNPDSLSEAMIDMLRKVKVSTSGEHITLFLHPILL 419
 QY 352 STNY 355
 Db 420 GYNH 423

RESULT 4
 Q9VSM4 PRELIMINARY: PRT: 436 AA.

ID Q9VSM4
 AC Q9VSM4
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, last sequence update)
 DT 01-MAR-2001 (TREMblrel. 16, last annotation update)
 DE ANON-66DB PROTEIN.
 GN CG5989 OR ANON-66DB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA MEDLINE=20196006; Pubmed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Chame M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasly E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
 RA Burlas K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriere S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalili M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mervulov G., Mitsuhashi N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon R., Nusskern D.R., Paclet J.M.,
 RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenklamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA "The genome sequence of *Drosophila melanogaster*."
 RT Science 287:2185-2195(2000).
 RL EMBL: AE003555; AAF50392.1;
 DR Flybase: FBgn0017429; CG5989.
 SQ SEQUENCE 436 AA: 50314 MW: 282EFDA91ACC5ADD CRC64:

Query Match 17.9%; Score 342; DB 5; Length 436;
 Best Local Similarity 26.6%; Pred. No. 1.3e-23;
 Matches 81; Conservative 67; Mismatches 148; Indels 8; Gaps 4;

OY 55 NLMSTVYTKAINGKYHFLGRHPPRYLYTIFMKGLQIMLADAKKARIK--TNMK 112
 DB 125 NMQDYIFTRYFNKYNDVLEKNPKAMQLYRFEDGKDFGDMKRLKARIANDSP 184
 OY 113 HNKFHQLYEMHNLQFRODVTKCLFLGITSIPFANVLYFLLMYLPROLLIRHEWT 172
 DB 185 QGIR--ALNRQELLYMQMPRDMKVPALIGCSLPWVGAFPLVYYPRESFLTAHEWT 242
 OY 173 PKQOTDFLDIYAFRKOSHPEITISYLEKVIPLISDAGLRMLRPLDCTKIORGTHPAIHDI 232
 DB 243 PQORSEFQSYMKRRLCNKNDVFRCLQDKATASHPKHSAPADILQSGSTHPTPEML 302
 OY 233 LALRECFSNHPLGMMNOLOALHVKALSRAMLITFSLPPLRLRHLKTHVTYIHOLOKALAK 292
 DB 303 IDVKDIFAECPYSILGMSRKHVRLNVLHGLPSSI--FKRHRLHEHAFVHYWDAITR 359
 OY 293 L-GIGQLTQAEVKSACYLRLGSLNTHIGEDRCRTWLGEMLQISCLKEALSLILHNVLL 351
 DB 360 EGVGHNLTPDALRYSCYLRGLNPDLSSEAMIDWLKRWKVSISIOGBHITFLFHLPIIL 419
 OY 352 STNY 355
 DB 420 GYNH 423

RESULT 5
 ID P91616 PRELIMINARY; PRT; 436 AA.
 AC P91616;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
 DE ANON-66DB.
 GN CG5989 OR ANON-66DB.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP STRAIN=CANTON S (WILD TYPE); TISSUE=WHOLE PUPAE;
 RX MEDLINE=97379971; PubMed=9236770;
 RA Saeboe-Larsen S., Urbanczyk Mohebi B., Lambertson A.,
 RT "The *Drosophila* ribosomal protein L14-encoding gene, identified by a
 RT novel Minute mutation in a dense cluster of previously undescribed
 RT genes in cytogenetic region 66D.";

RL Mol. Gen. Genet. 255:141-151(1997).
 DR EMBL: Y10018; CAA71125.1;
 DR Flybase: FBgn0017429; CG5989.
 SQ SEQUENCE 436 AA: 50200 MW: F0D5EC0B5BAB00F6 CRC64:

Query Match 17.8%; Score 341; DB 5; Length 436;
 Best Local Similarity 26.6%; Pred. No. 1.7e-23;
 Matches 81; Conservative 67; Mismatches 148; Indels 8; Gaps 4;

OY 55 NLMSTVYTKAINGKYHFLGRHPPRYLYTIFMKGLQIMLADAKKARIK--TNMK 112
 DB 125 NMQDYIFTRYFNKYNDVLEKNPKAMQLYRFEDGKDFGDMKRLKARIANDSP 184
 OY 113 HNKFHQLYEMHNLQFRODVTKCLFLGITSIPFANVLYFLLMYLPROLLIRHEWT 172
 DB 185 QGIR--ALNRQELLYMQMPRDMKVPALIGCSLPWVGAFPLVYYPRESFLTAHEWT 242
 OY 173 PKQOTDFLDIYAFRKOSHPEITISYLEKVIPLISDAGLRMLRPLDCTKIORGTHPAIHDI 232
 DB 243 PQORSEFQSYMKRRLCNKNDVFRCLQDKATASHPKHSAPADILQSGSTHPTPEML 302
 OY 233 LALRECFSNHPLGMMNOLOALHVKALSRAMLITFSLPPLRLRHLKTHVTYIHOLOKALAK 292
 DB 303 IDVKDIFAECPYSILGMSRKHVRLNVLHGLPSSI--FKRHRLHEHAFVHYWDAITR 359
 OY 293 L-GIGQLTQAEVKSACYLRLGSLNTHIGEDRCRTWLGEMLQISCLKEALSLILHNVLL 351
 DB 360 EGVGHNLTPDALRYSCYLRGLNPDLSSEAMIDWLKRWKVSISIOGBHITFLFHLPIIL 419
 OY 352 STNY 355
 DB 420 GYNH 423

RESULT 6
 ID Q25474 PRELIMINARY; PRT; 364 AA.
 AC Q25474;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-NOV-1998 (TREMBLrel. 08, Last annotation update)
 DE LEUCINE ZIPPER PROTEIN.
 GN URO-2.
 OS *Molgula oculata* (Sea squirt).
 OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea;
 OC Stolidobranchia; Molgulidae; Molgula.
 OX NCBI_TaxID=27575;
 RN [1]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=94116436; PubMed=8287790;
 RA Swalla B.J., Makabe K.W., Satoh N., Jeffery W.R.,
 RT "Novel genes expressed differentially in ascidians with alternate
 RT modes of development."
 RT Development 119:307-318(1993).
 DR EMBL: L19340; AAC37181.1;
 SQ SEQUENCE 364 AA: 42609 MW: D38DA95B30AA04E CRC64:

Query Match 10.8%; Score 206; DB 5; Length 364;
 Best Local Similarity 22.5%; Pred. No. 4.9e-11;
 Matches 81; Conservative 64; Mismatches 141; Indels 74; Gaps 9;

OY 46 HUSPADKKNLMSTVYTKAINGKYHFLGRHPPRYLYTIFMKGLQIMLADAKKARIK--TNMK 114
 DB 26 HUSPADKKNLMSTVYTKAINGKYHFLGRHPPRYLYTIFMKGLQIMLADAKKARIK--TNMK 114
 OY 91 KGIOM-----LADAKKARIKTNMKHINIKFQLPYREHNLQFRODVTKCLFLGII 144
 DB 86 AGLRQOKQFRYWRSAOKKKIDPHSY-----ITRREKALITLLKKDKFLKATMI--- 134
 OY 145 SIP-----PFANVLYFLLMYLPROLLIRHEWTYPQOQDFLDIYAFRKOSHPEITISYLE 199
 DB 145 SIP-----PFANVLYFLLMYLPROLLIRHEWTYPQOQDFLDIYAFRKOSHPEITISYLE 199

```

Db 135 -VPTFLVPGFLFLIPEVLPYVLPDCYWRDQSKYLTQATSRAYGMILHLS 193
OY 200 -KVPLISDAGLRWRLDCTKIORGTHPAIHDLALRECFSNHPLGMNOALHVKALS 258
Db 194 YHKNLISODRAORLHALVOQVSEGVPSYDLSLHLPLFRD-----S 237
OY 259 RAMELTSYLPPLLRHRLKHTTV-----IHOLDKALAKG-IGOLT 299
Db 238 SSLRLISDLNPNVRLSCNTAFIVPFOTKSMALRSLTRMDLFLRLDRKLRDEDLHKIS 297
OY 300 AOEKVSACVLRGTHIGEDRCRTWLGEMLOISCSLKEAEELSLHANYLSTNLGTR 359
Db 298 PLHLEATYLRKGLDLSLSPDANKRFLQHWLQVTRNCSDUDTVFMHAWLMSFNSYSEK 357

RESULT 7
095202 PRELIMINARY: PRT: 739 AA.
AC 095202;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LEUCINE ZIPPER-EF-HAND CONTAINING TRANSMEMBRANE PROTEIN 1.
GN LETM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Endele S., Pak S.-J., Fuhry M., Zabel B., Winterpacht A.;
RT "LETM1, a novel leucine zipper and EF-hand containing transmembrane
RT protein gene located on chromosome 4p16.3."
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: NO EF-HAND FAMILY.
EMBL: AF061025; AADI3138.1;
DR InterPro: IPR002048; EF-hand.
DR Pfam: PF00036; efhand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Calcium-binding; Transmembrane.
SQ SEQUENCE 739 AA; 83353 MW; 942E9138F29D94F CRC64;

Query Match 9.4%; Score 180.5; DB 4; Length 739;
Best Local Similarity 23.8%; Pred. No. 2.8e-08;
Matches 88; Conservative 60; Mismatches 160; Indels 61; Gaps 14;

OY 3 LSRVNCARSAVNGSAVTPRGHFTYRQLQGR-----SGLAWGPRSSKHL-----LS 48
Db 83 VSRAPWTSTVCGVAVAGPQCLPYRGHSSRPVRRDSVVERKSLSKDKKKLEEGGPVYS 142
OY 49 PKADV--KNLSYVVTKTALNGKYHRELGHHFPRFYTLTYIFMKGLOMLWADAKKAR 105
Db 143 PRAEVVVKKSLGQRLVDLKHV---YH-----GRLMLMDITKIAAR 180
OY 106 IKTNNMKHNKIFQOLPYREMEHLKQFRODYTCFLGITSIPPEANVLYFLMLYFPROL 165
Db 181 M--LWR-ILNGISLTRRRROFLRICADLFRVLPFLVEVVPFMFLPVAVKLF--NM 235
OY 166 LIRHFWTPROQDFLDIYHAFRKOSH--EISYLEKVI---PLISDAGLRRLDLDLCK 220
Db 236 LSTFTFQSLKEERL-----KELRVKLELAKFLQDTLEEMALKKAKKAGSATDFSVF 289
OY 221 IGR---GTHPAIHDLALRECFSNHPLGMNOALHVKALSRAMLTSYLPPLLRHL 276
Db 290 FOKIRFTGRSPSNEEIMRSKLFEDL-LTLDNLTRPQVALCKLELDSIGTNNLRQL 348
OY 277 KHTTYIHDLAKLAKLGIGQLTAQEVKACVLRGINSTHIGEDRCRTWLGEMLOISCSL 336
Db 349 TMRLSIKADDLIAEEGVDSINVELQAACRARGMALGVTEDRQLKQMDLHLH- 407
OY 337 KEAELSLLL 345

```

```

Db 408 QEIPTSLIL 416

RESULT 8
092210 PRELIMINARY: PRT: 738 AA.
AC 092210;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LEUCINE ZIPPER-EF-HAND CONTAINING TRANSMEMBRANE PROTEIN 1.
GN LETM1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Endele S., Pak S.-J., Fuhry M., Zabel B., Winterpacht A.;
RT Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF061026; AADI3139.1;
DR InterPro: IPR002048; EF-hand.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
KW Transmembrane.
SQ SEQUENCE 738 AA; 82988 MW; 5074CE630940729 CRC64;

Query Match 8.9%; Score 170; DB 11; Length 738;
Best Local Similarity 24.4%; Pred. No. 2.6e-07;
Matches 65; Conservative 54; Mismatches 125; Indels 22; Gaps 9;

OY 89 FMKGLOMLWADAKKARRTKTNMKHNKIFQOLPYREMEHLKQFRODYTCFLGITSIP 148
Db 163 YHNGRLLMDITKIAAR--LWR-ILNGHTLTRRRROFLRICADLFRVLPFLVAVVP 218
OY 149 FANYLVFLMLYFPRQLIRHFWTPKQOTDFLDIYHAFRKOSH--EISYLEKVI---P 203
Db 219 FMEFLPVAVKLF--NMLPSTFETQSLKEERL-----KELRVKLELAKFLQDTLEMA 271
OY 204 LISDAGLRRLDLDLCKIQR---GTHPAIHDLALRECFSNHPLGMNOALHVKALSR 259
Db 272 LKNKAAGKMATKDSAFQKIRETGERSPNEIMRFSKLFEDL-LTLDNLTRPQVALCK 330
OY 260 AMLTSYLPPLLRHRLKHTTYIHOLDKALAKLGIGQLTAQEVKACVLRGINSTHIGE 319
Db 331 LLELQSTITNPFRLKQQLMRLSIKADDKLISEGVSLSLYKELQAACRARGMALGVTE 390
OY 320 DRCRTWLGEMLOISCSKEAELSLLL 345
Db 391 DRLKQQLKQMDLHLH-HEIPTSLIL 415

RESULT 9
004471 PRELIMINARY: PRT: 398 AA.
AC 004471;
DT 01-JUL-1997 (TREMBLrel. 04, Created)
DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE SIMILAR TO SACCCHAROMYCES HYPOTHETICAL PROTEIN P9642.2.
GN F5114.7.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC eucotsids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Vysotskaia V.S., Osborne B.I., Toriumi M., Yu G., Oji O., Shen Y.K.,
RA Buehler E., Conway A.B., Conway A.R., Dewar K., Feng J., Kim C.,
RA Kurtz D., Li Y., Sun H., Davis R.W., Ecker J.R.,
RA Federspiel N.A., Theologis A.;

```


QY	287	DNAL-AKLIIGOLTQOEKASACVYGLNSTHIGEDRCRTYGEMLQYSCSLAKEAEISLL	343
Db	299	DKQIAAEGGVDAISLSDIQSACRANGRAIGVSEERKEQLVQWLESLNDKVPALLL	358
QY	346	HNWVLL 351	
Db	359	SRTLYL 364	
RESULT	12		
Q9XVMO	Q9XVMO	PRELIMINARY;	PTT; 784 AA.
AC	Q9XVMO;		
DT	01-NOV-1999	(TREMBLrel. 12, Created)	
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)	
DT	01-MAY-2000	(TREMBLrel. 13, Last annotation update)	
DE	F58611.1A	PROTEIN.	
GN	F58611.1A.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;		
OC	Rhabditidae; Pelodierinae; Caenorhabditis.		
OX	NCBI_TaxID=6239;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RL	percy C.;		
RL	Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RL	MEDLINE=94150718; Pubmed=7906398;		
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,		
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,		
RA	Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,		
RA	Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,		
RA	Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,		
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.		
RA	Smaldon N., Smith A., Sonhammer E., Straden R., Sulston J.,		
RA	Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,		
RA	Watson A., Weinstock L., Wilkinson-Sprout J., Wohlman P.;		
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
RT	elegans".		
RL	Nature 366:32-38(1994).		
DR	EMBL; Z81094; CAB03150.1; .		
SO	SEQUENCE 784 AA; 88721 MW; B5421FF51EA6E9F CRC64;		
Query Match	7.38; Score 140; DB 5; Length 784;		
Best Local Similarity	23.28; Pred. No. 0.00018;		
Matches 57; Conservative 52; Mismatches 113; Indels 24; Gaps			
QY	120	LPYREMEHLRFODVYKCLFLGIISIPPNAYLVFLMYLFPROLLIRHFWPKQOTDF	179
Db	129	LSRRRQQLVATVSDRLRVFSEFTITVPPMELALPFIKLFPGMLPSFFQSSKEE---	185
QY	180	LDIYHAERKQ--SHEPILSYLEKVIPLI-----SDAGLRWRRLDTCTKIQ-RGTH	226
Db	186	----EKMKQKQYLAREVMAKFLQDTIEIEIGLERKTRNKSETRSLFPL--FIKVRNEGQY	239
QY	227	PAIHDIILARCFSENHPLCGMNOIALHVKALSRAMLLTSTLPPLLRHRKTHHTYHOL	286
Db	240	VSNEELLKFSKLFDE--ITLDNLSMGQLRSKRLMSINSIGSPETILRFOLDNMKIRELKAD	298
QY	287	DKAL-AKLIIGOLTQAOEVKASACVYGLNSTHIGEDRCRTYGLGWLQISLKEAEISLL	345
Db	299	DKQIAAEGGVDAISLSDIQSACRANGRAIGVSEERKEQLVQWLESLNDKVPALLL	358
QY	346	HNWVLL 351	
Db	359	SRTLYL 364	
RESULT	13		

ID	09M122	PRELIMINARY:	PRT:	755 AA.
AC	09M122:			
DT	01-OCT-2000 (TReMBLrel. 15, Created)			
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)			
DT	01-JUN-2001 (TReMBLrel. 17, Last annotation update)			
DE	HYPOTHETICAL. 85.7 KDA PROTEIN.			
F24G16.90.				
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	D'Angelo M., Vezzi A., Modesto D., Pigazzi M., Valle G., Mewes H.W.,			
RA	Lemcke K., Mayer K.F.X., Queller F., Salanoubat M.,			
RL	Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;			
RL	Submitted (Feb-2000) to the EMBL/GenBank/DBJ databases.			
CC	-I- SIMILARITY: TO EF-HAND FAMILY.			
DR	EMBL: AL138647; CAB75801.1; -			
DR	InterPro: IPR002048; EF-hand.			
DR	Pfam: PF00036; ehand; 1.			
DR	PROSITE: PS00018; EF-HAND; UNKNOWN_1.			
KW	Calcium-binding; Hypothetical protein.			
SQ	SEQUENCE 755 AA; 85734 MW; F631EP9307AF6108 CRC64;			
Query Match	7.1%;	Score 136.5;	DB 10;	Length 755;
Best Local Similarity	19.7%;	Pred. No. 0.00036;		
Matches	50;	Conservative 56;	Mismatches 123;	Indels 25;
Gaps				
QY	92 GLQMLADAKKRRRIKTNMKNNHINFEHQHPYEMEHLLROFQDVTKCFILGIIISIPFAN 151			
DB	218 GFKLLMADPTRISSRLKLT---AGSKLSRREKQRLRTTADIRLVFVAFILVPPNE 273			
QY	152 YIVFLMLLFPQQLLRHFWTPKQOTDFLDIYHAFKQ--SHPEIISYLEKVIPLISDAG 209			
DB	274 FLPLPFLKLFPMNL-----PSTPDKKKEEALRKLLAIEYAKFLEQETAREMAKEV 326			
QY	210 LKWR-----LTDCTKLRGCTGTHPAIHIIILRECFNHPHGMNOLQALHKAISR 259			
DB	327 KHSRFGVKKQIADDEDFDLKVRKQOIYHNDELLFAKLF-NDELITDINSRPLVSCK 385			
QY	260 AMLTSTYLPPLLRHRLKTHVTIVHQLDKALAKIGITLQAEVNSACYLRGLNSTHGE 319			
DB	386 YMGISPYGDAVLRMYLKRKRKRSIKEDDKLIRAEVSDLSSEALREDCRERGMGLT- VSV 444			
QY	320 DRCPTLGMELQIS 333			
DB	445 EEMQQLRDMWDL 458			
RESULT	14			
008179				
AC	008179:	PRELIMINARY:	PRT:	573 AA.
DT	01-NOV-1996 (TReMBLrel. 01, Created)			
DT	01-NOV-1996 (TReMBLrel. 01, Last sequence update)			
DT	01-MAR-2001 (TReMBLrel. 16, Last annotation update)			
DE	CHROMOSOME XV READING FRAME ORF YOL027C.			
YOL027C.				
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyc.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Habib B., Hattenhorst U., Hollenberg C.P., Ramezani Rad M.,			
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.			

RN [2]
 RP SEQUENCE FROM N.A.
 RA MIPS;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Z74769; CA99027.1; -
 DR SCD; S0005387; YOLO27C.
 SQ SEQUENCE 573 AA; 65005 MW; 5F34453EC09F311 CRC64;

Query Match 7.1%; Score 136; DB 3; Length 573;
 Best Local Similarity 23.2%; Pred. No. 0.00028;
 Matches 68; Conservative 58; Mismatches 117; Indels 50; Gaps 12;

QY 89 FMKGLQMLMDAKKARRIKTNMKKNIKF---HQLPYREMEHLROFRODVTCLEGLIIS 145
 DB 94 YANGTKLGEIKVSTKFL-----IKFAGVELSRNRNOLKRTMGVDVRLIFPSAF 146
 QY 146 IPEFANVYFLMYLFP-----ROLLIRHFWTPKQOQDFDIYHAFKQ 189
 DB 147 IIPFALFELPFALKLPNLLPSTYESGKDOKARKNKLIE--IRKKTSEFL--HETLEE 200
 QY 190 SHPEITISYLEKYIPLISDAGLRMLTDLCTKIQGTGHPAI---HD-LLALRECSNHP 244
 DB 201 SN--LITY-----NTIENKCKKFLNFRKIKYSAKEGKIMTFQHDSTSAIAQMFKNDSV 253
 QY 245 GMMQLOALHWKALSRAMILTLSTLPPLLRHRLKTHTVIHOLDKALAKIGIGOLTAQEVK 304
 DB 254 -LDNISRPOLAMSKFMSLRPFQNDNMLRYOIRSKLKDIMNDKTIYEGVESLSQEE 312
 QY 305 SACYLNGLSNTHIGEDRCRTWIGEWLQISCSLKEALSLILAHNVLLSTNIG 357
 DB 313 QACVSRGMKMGVSKEDLVNDLKVLEL--RLRQKIPSVLM--VLSSTFEG 360

RESULT 15

Q93654 PRELIMINARY; PRT; 497 AA.

AC Q93654;
 DT 01-FEB-1997 (Tremblrel. 02, Created)
 DT 01-FEB-1997 (Tremblrel. 02, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE F30F8.3 PROTEIN.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;

[1]
 RP SEQUENCE FROM N.A.
 RA Wilkinson J.;

RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;

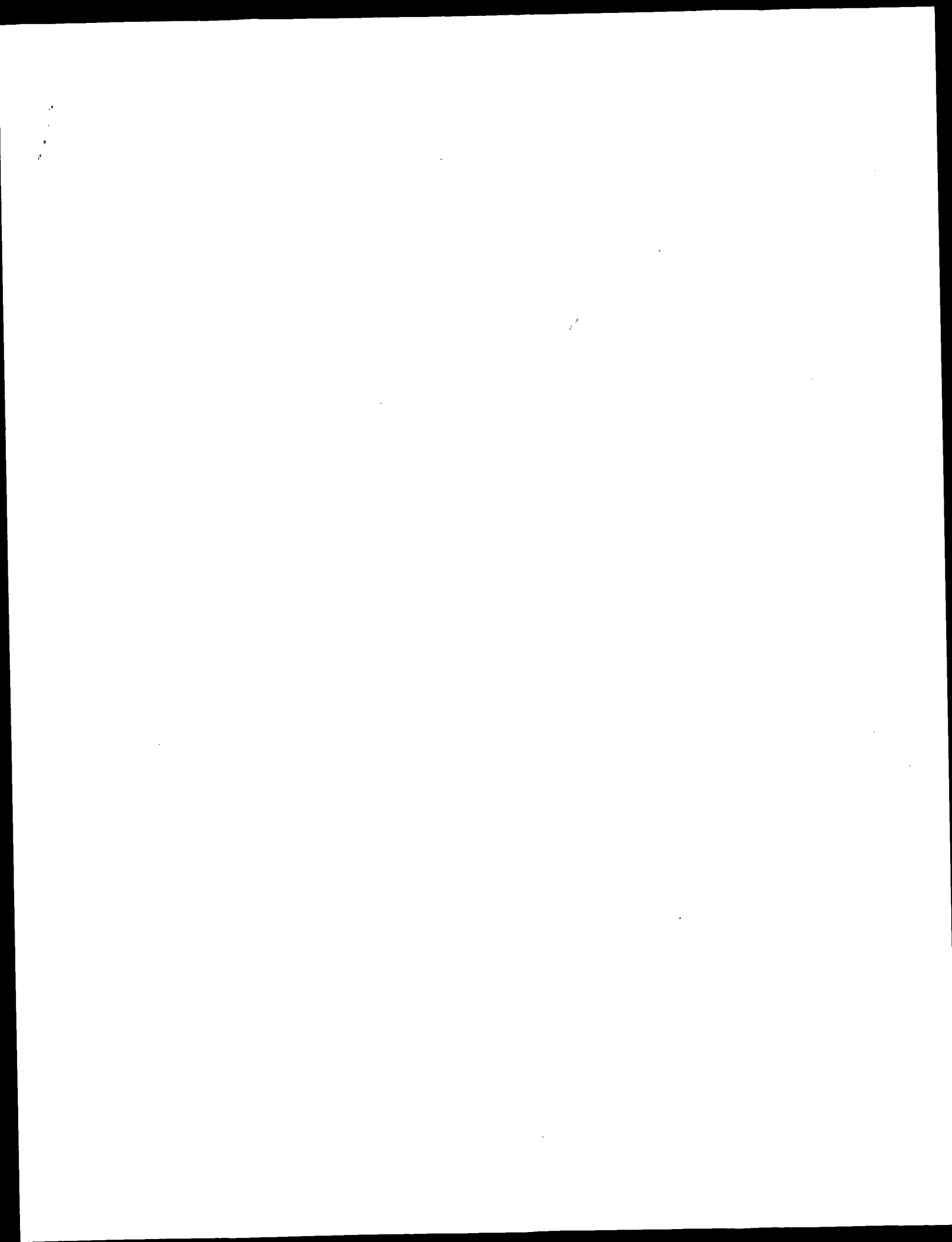
RA Wilson R., Alinscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Koopa A., Saunders D., Showkneen R.,
 RA Smaison N., Smith A., Sonnenhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Woldman P.,
 RT 2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.;

RL Nature 368:32-38(1994).
 DR EMBL: Z81073; CA903034.1; -
 DR InterPro: IPR001478; PDZ.
 DR Pfam: PF00595; PDZ; 1.
 DR PROSITE: PSS0106; PDZ; 1.
 DR SMART: SM00228; PDZ; 1.
 SQ SEQUENCE 497 AA; 56936 MW; B97BD20A83DF9468 CRC64;

Query Match 6.9%; Score 131.5; DB 5; Length 497;
 Best Local Similarity 21.0%; Pred. No. 0.00062;
 Matches 62; Conservative 51; Mismatches 127; Indels 55; Gaps 8;

QY 63 KTKAINGYHRFLGRHPREFYLYITFMKGLQMLMDAKKARRIKTNMKKNIKFHOLPY 122
 DB 248 KESGLTYKIQFQIGK-WPKYLAHRTVLGSRKCEADVMYFKTKRAVATGOKLTLDSV 306
 QY 123 REMEHLRQFRODVTKCLFGIISIPFPANYVFLMYLTPROLIRHFWTPKQOQDFDI 182
 DB 307 EELETLVQMPVEGPKMAIVTAFLPPLSYVFAFALIEPRLVLIRHFWSDQQRREYQ- 365
 QY 183 YHAFKQSHPEITISYLEKYIPLISDAGLRMLTDLCTKIQGTGHPAIDILALRECSNH 242
 DB 366 -----LEVTKALISGEOL-----LKTGNGPSSYD----- 389
 QY 243 PLGMNOLALHWKALSRAMILTLSTLPPLLRHRLKTHTVIHOLDKALAKIGIGOL 298
 DB 390 ---ENKMKPLD-KIDSEMLLVHNMQSMYPIPGSSKRIMNKKALRALDNLPTT-IDGF 444
 QY 299 TAOEVKSACYLNGLSNTHIGEDRCRTWIGEWLQISCSLKEALSLILAHNVLLST 353
 DB 445 NEQLIFNCYIKRIDIGKKNSEMRDSLROYVKF-----SRMPNVLYLAS 491

Search completed: December 20, 2001, 10:12:31
 Job time: 119 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 16:09:03 ; Search time 1816.46 Seconds

(without alignments)
12529.623 Million cell updates/sec

Title: US-09-868-474-1

Perfect score: 2118

Sequence: 1 ctgtgaagatgagctctcc.....aaaaaaaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 segs, 537289281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estlin:*

4: em_estlom:*

5: em_estlpl:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_hic:*

10: gp_estl:*

11: gp_est2:*

12: gp_hic:*

13: gp_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_tod:*

20: em_gss_vtl:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	913.6	43.1	951	10	AL542968
C 2	848.2	40.0	885	10	AL576116
C 3	784.4	37.0	864	10	AL576116
C 4	772.6	35.5	801	10	AL576271
5	757.4	35.8	863	11	AL579274
6	738.2	34.9	849	11	AL579274
7	736.6	34.8	819	11	AL579274
8	731.4	34.5	804	11	AL579274
9	717.4	33.9	804	11	AL579274
10	713.4	33.7	850	11	AL579274
11	711.8	33.6	844	11	AL579274
12	700.2	33.1	967	11	AL579274

13	689.4	32.5	702	11	BI255100
14	686	32.4	714	11	BI255100
15	681	32.2	702	10	BI255100
16	672.4	31.7	687	11	BI255100
17	664	31.4	740	11	BI255100
18	658.2	31.1	694	10	BI255100
19	655	30.9	876	10	BI255100
20	649.8	30.7	889	11	BI255100
21	642.8	30.3	1085	11	BI255100
22	635.2	30.0	672	10	BI255100
23	634	29.9	827	11	BI255100
24	617.8	29.2	618	11	BI255100
25	606.4	28.6	641	11	BI255100
26	604.4	28.5	615	10	BI255100
27	597	28.2	713	10	BI255100
28	591.4	27.9	601	10	BI255100
29	576	27.2	649	10	BI255100
30	571.4	27.0	790	11	BI255100
31	558	26.3	605	10	BI255100
32	550	26.0	665	10	BI255100
33	548.4	25.9	1111	10	BI255100
34	542.4	25.6	548	10	BI255100
35	539.8	25.5	748	11	BI255100
36	539.4	25.5	561	10	BI255100
37	534.8	25.3	591	11	BI255100
38	532.2	25.1	595	11	BI255100
39	530.6	25.1	791	10	BI255100
40	523.8	24.7	725	11	BI255100
41	522.6	24.7	731	11	BI255100
42	520.8	24.6	795	10	BI255100
43	518	24.5	529	10	BI255100
44	518	24.5	946	11	BI255100
45	517.4	24.4	530	10	BI255100

ALIGNMENTS

RESULT 1	AL542968	951 bp	MRNA	EST	16-FEB-2001
LOCUS	AL542968	951 bp	MRNA	EST	16-FEB-2001
DEFINITION	AL542968	951 bp	MRNA	EST	16-FEB-2001
ACCESSION	AL542968	951 bp	MRNA	EST	16-FEB-2001
VERSION	AL542968	951 bp	MRNA	EST	16-FEB-2001
KEYWORDS	AL542968	951 bp	MRNA	EST	16-FEB-2001
SOURCE	AL542968	951 bp	MRNA	EST	16-FEB-2001
ORGANISM	AL542968	951 bp	MRNA	EST	16-FEB-2001
REFERENCE	AL542968	951 bp	MRNA	EST	16-FEB-2001
AUTHORS	AL542968	951 bp	MRNA	EST	16-FEB-2001
TITLE	AL542968	951 bp	MRNA	EST	16-FEB-2001
JOURNAL	AL542968	951 bp	MRNA	EST	16-FEB-2001
COMMENT	AL542968	951 bp	MRNA	EST	16-FEB-2001
FEATURES	AL542968	951 bp	MRNA	EST	16-FEB-2001
Source	AL542968	951 bp	MRNA	EST	16-FEB-2001

Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web: www.genoscope.cns.fr.

Location/Qualifiers

1. 951

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="CS0DE013YH04"

/clone_id="LTI_FL002_PU1"

/lab_host="DH10B"

/note="Organ: Placenta; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with NotI and cloned into the NotI and EcoRV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact: Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax: (1) 301 610 8371 Email: "

filang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 299 a 222 c 201 g 211 t 18 others
ORIGIN

Query Match 43.1%; Score 913.6; DB 10; Length 951;
Best Local Similarity 96.5%; Pred. No. 4.5e-148;
Matches 913; Conservative 17; Mismatches 16; Indels 0; Gaps 0;

1102 gaacgagatgcatgtccgcagctgcatagatagacagtcaggaacaacagacatg 1161
1102 gaacgagatgcatgtccgcagctgcatagatagacagtcaggaacaacagacatg 1161
951 GACCGATGGCATGTCTCGACAGTCTGATATACAGTCAAGCAACAGACAGCTTG 892
1162 ccagcaagtcgtgtgtctgttaagtggtgagagagagagagagagagagagat 1221
891 CCAGCAAGTCTGTGTCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 832
1222 gggcttcacagcagtcagcagcagtcagcagtcagcagtcagcagtcagcagtc 1281
831 GGGCTTCACAGCAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 772
1282 aaacaaaccctctgctagagagtggtgcgltgagagtgatcctgtccctcataat 1341
771 AAACAAACCCTTGTGCTAGGGGTGTCCGTGTGAGTGTGATCTGTCCCTCATTAAT 712
1342 actaatagctggaactggaagcagcagcagcagcagcagcagcagcagcagcagc 1401
711 ACTAATAGCTGGAAGTGGAGCAGCCTTACGTGGCTTTACTGTATGTATGTGTGTGT 652
1402 tgcctcaggaagtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1461
651 TGCTCCTAGGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 592
1462 actcagatgcatcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1521
591 ACTTCCATGTACATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 532
1522 tgaagaaagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1581
531 TGAAGAAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 472
1582 tcaactcctcagatagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1641
471 YCCATCTCTTCTGATGATGACCAAAATCAAGTGTGTGTGTGTGTGTGTGTGTGTGT 412
1642 atggagtaacacactctcttcaaacctctgagacactcttctcagtgagtaactg 1701
411 ATGGGCTAACCACTTCTTTTCAAAACCTTGAAACACCTTTTCTGTGTGTGTGTGTG 352
1702 ggaatattcattggaagaagaagaagaagaagaagaagaagaagaagaagaagaaga 1761
351 GGAATATTCTTCTTGGAAAAGATTAACAGAGTACAACTGCTTGTGACCCCTTCTCAAT 292
1762 gtctcagcctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1821
291 GTTCTAGCCTTCACTCTCCATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 232
1822 tcaactcgtcgtcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1881
231 TCAATTTCTGTGCTCTCCAGCTGTGATGAGTGTGTGTGTGTGTGTGTGTGTGTGTGT 172
1882 ctctggagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1941
171 TTTTGGGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 112
1942 agaggtcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 2001
111 AGAGTCTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 52
2002 tctcgtataactcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2047
51 TTTTATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 6

RESULT 2
AL576116/c 885 bp mRNA EST 16-FEB-2001
LOCUS
DEFINITION AL576116 L11_NFL006.PL2 Homo sapiens cDNA clone CSDD10721B20 3
ACCESSION
VERSION AL576116
KEYWORDS AL576116.1 GI:12937941
SOURCE
ORGANISM human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 885)
Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
Full-length cDNA libraries and normalization
unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Source
Location/Qualifiers
1..885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CSDD10721B20"
/clone_lib="L11_NFL006.PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : filang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 265 a 208 c 204 g 194 t 14 others
ORIGIN

Query Match 40.0%; Score 848.2; DB 10; Length 885;
Best Local Similarity 97.2%; Pred. No. 8.6e-137;
Matches 860; Conservative 12; Mismatches 12; Indels 1; Gaps 1;

1118 cctgagctgctagatagcagtcagcagcagcagcagcagcagcagcagcagcagcagc 1177
885 CCGCAGCTCTTATGATATGACAGTGCAGGACAAACAGCAGCTGCGACAGCAAGCTGTGT 826
1178 gtaactgtaagtggtgagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1237
825 GTAAGTGTAAAGT 766
1238 caacactgaggagactgagcagcagcagcagcagcagcagcagcagcagcagcagc 1297
765 CACACATGTGGGAAGTGCAGACATCCCTCTCAAGTGAAGTGAACCAACCTCTTCTGC 706
1298 taagggtgctcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1357
705 TAGGGGTGTGCTGT 646
1358 ggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 1417
645 GGCAGCAGCAGCTTACAGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 586
1418 cttctg-cgccagtggtgagcagcagcagcagcagcagcagcagcagcagcagcagc 1476
585 CTTTGTCCCGCCAGCTGGAGATCTTATTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 526
1477 acatcgtgaggtttgt 1536

LOCUS	AL579274	801 bp	mrna	EST	16-FEB-2001
DEFINITION	AL579274 LTL_FL012.TC1 Homo sapiens cDNA clone CS0DH002J714 3 prime				
ACCESSION	AL579274				
VERSION	AL579274.1				
KEYWORDS	GI:12944155				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr. location/qualifiers				
FEATURES	<p>source</p> <p>1. 801</p> <p>/organism="Homo sapiens"</p> <p>/db_xref="taxon:9606"</p> <p>/clone="CS0DH002J714"</p> <p>/clone_1bp="LTL_FL012.TC1"</p> <p>/tissue_type="T cells from T cell leukemia"</p> <p>/lab_host="DH10B"</p> <p>/note="Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by lile technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"</p>				
BASE COUNT	232 a 192 c 190 g 178 t 9 others				
ORIGIN					
Query Match	36.5%; Score 772.6; DB 10; Length 801;				
Best Local Similarity	97.8%; Pred. No. 9.5e-124;				
Matches 783; Conservative	9; Mismatches 8; Indels 1; Gaps 1;				
OY	1080	acaagcgctgaatgaaccatgagcgatgcatgtctctgcagtcgtatagatagca	1139		
Db	801	ACAAGGCCCTGATTAACCATGAGACGGATGCAATTCCTGCACTGATATGATACGA	742		
OY	1140	gtgcagagacaacagcgacttcgcagcaagctgtgtactgttlaagtgtgtggaag	1199		
Db	741	GTGCGAGGACAAACAGCACTTCCAGCAAGCTGTGTGTACTGTTAAGTGTGTGGAGC	682		
OY	1200	cagaagagagagcaagggccatgggcttcacagcaltgagacacctgtggaactgcagac	1259		
Db	681	CAGAGAGAGGAGACGAGGGCCATGGGCTTCACAGCATGGCACACATGTGGGCACTGCAGAC	622		
OY	1260	attcctc-tcaagctagaactgaaacaaaccttcttaggggtgttcgttgaagt	1318		
Db	621	ATTCTCTTTCACAGGTAGAAATGAACAACCCCTTGCTAGGGGTGCTCCGTGTGAGGT	562		
OY	1319	gtcacctcgtcccccataattactaatagcttgaaactgcagacagcctcactcgggt	1378		
Db	561	GTCATCTCTGTCCTCCTCAATTAAGTGTGAACGAGGACGACGCTTACTAGGGGT	502		
OY	1379	ttactgtgatgtgttcagttcatcgttcctaggaatcagcttltgcccagtggaatc	1438		
Db	501	TTTACTGTGATGTGTTTCAGTTTCATGTCTCCAGGAGATCACTTTTCCCGCAGGTGGGAATC	442		
OY	1439	cttatttgcttaagactgaactcacttccatgttactctaacctctgtaggtttttctgt	1498		
Db	441	CTTATTTTGGCTTAGGACTGATCCACTTCATGTATCTACATCTGTGGGTTTTTTGTTGT	382		
OY	1499	gctgtttagaanaattlttgctgtttaaacaagacaccccttgcctgctggaagacctgtgtcc	1558		

Db	381	GCATTGAGAAATTGTTGGGCTGGTGAAACACAGACACTCCTTTGGCTGAGACACTTGTCGCC	322
Qy	1559	atgcagatgaacttggtggtttccctccatccctcttcgtatgatgacccaataaagtgtttc	1618
Db	321	GTGATGTTACTTGGTGGTGGTTCCTCCATCTCTTGTGATATGACCAAAATCAAGTTGGTT	262
Qy	1619	tgtttttgtccacttcaactgcatgcatgagcttaacacacttcttttcaaacctctgaaac	1678
Db	261	TGTTTTTTTGTACACTTCACTGAGATGGGCTAACCAACTTCTTTKTCAAACCTTGACAC	202
Qy	1679	ctttttctgattggttaactgtgcaggaataattcttatgtgaaaaagataacaggaagtacaag	1738
Db	201	CTTTTCTGTATGGGTAACTGGCAGGAATATCTATTGGAAAAAGATAAACAGAACTAACAAG	142
Qy	1739	tgcctcttgagcccttcctcaatggtttctcagacttcaacttccatctgtctttcttgagact	1798
Db	141	TGCTTCTTGACCCCTTCCTCCTCAATGATTTTCAAGCTTCACCTTCACATGTCCTTTCTGGGCT	82
Qy	1799	gtattacagccctctgtgatcttcaactctgctgctcctcaactgtgatgcagcagccaa	1858
Db	81	GTATTACAGCCCTCTGGATCTTCAACTCTGCTCCTCCACTGATGAGKAGCAGACACAA	22
Qy	1859	ctgtaactgacagtgcgtgcgc	1879
Db	21	CTGTAAGTGAAGTGGCTGCC	1
RESULT	5		
LOCUS	BI085329	863 bp	mRNA
DEFINITION	602870389P1 NIH_MGC_98	Homo sapiens	CDNA clone IMAGE:5013406 5',
ACCESSION	BI085329		mRNA sequence.
VERSION	BI085329.1	GI:14503659	
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE	1 (bases 1 to 863)		
JOURNAL	NIH-MGC http://mhc.nhl.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgabos@email.nih.gov		
	Tissue Procurement: ATCC		
	CDNA Library Preparation: Ling Hong/Rubin Laboratory		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LNL at:		
	http://image.lnl.gov		
	Plate: LNCM819 row: a column: 23		
	High quality sequence stop: 814.		
	Location/Qualifiers		
	1. 863		
FEATURES	/organism="Homo sapiens"		
SOURCE	/db_xref="taxon:9606"		
	/clone="IMAGE:5013406"		
	/clone_lib="NIH_MGC_98"		
	/tissue_type="astrocytoma grade IV, cell line"		
	/lab_host="DH10B (phage-resistant)"		
	/note="Organ: brain; Vector: pOT7; Site:1: XhoI; Site:2:"		
	EcoRI; cDNA made by oligo-dT priming. Directionally		
	cloned into EcoRI/XhoI sites using the following 5		
	adaptor: GGCACGAG(G). Library constructed by Ling Hong		
	in the laboratory of Gerald M. Rubin (University of		
	California, Berkeley) using ZAP-cDNA synthesis kit		
	(Stratagene) and Superscript II RT (Life Technologies).		
	Note: This is a NIH_MGC library."		
BASE COUNT	212 a	226 c	192 g 233 t
ORIGIN			

Query Match	35.8%	Score 757.4	DB 11	Length 863
Best Local Similarity	96.9%	Prod. No. 3.8e-121		
Matches 825	Conservative 0	Mismatches 121	Indels 5	Gaps
QY	23	gggtgcgtggcctgcgtcgtcgtctgtgtgggctcggcaatcaccccttgacattctgcac	82	
Db	12	GGGTGCTGGGGCTGGCTGGCTGTGTGGGGGCTCGGAGTACCCCTTGACATTTTGTAC	71	
QY	83	ccggagagctgcaactgtgcgtctgcgttgcttgggtggggccctcgtcttcaagct	142	
Db	72	CCGGAGGCTGGCAACTGTGCTTGCCCTGGCTTGGGGGGGCCCTCGCTTCAAGCT	131	
QY	143	taacacctcccaaaagcagatgtgaagaacattgatactatgttgtaacaaagaaa	202	
Db	132	TCACCTTTCTCCAAAGGCGAGATGTAAAGACTGATGTATGTGTGTAAACGAACAA	191	
QY	203	agcgatbaatgagaaatacacatcgtcttcttggtgcgtcattcccccgctcatalcct	262	
Db	192	AGCGATTAATGGGAATATACATCGTTTCTTGGGTGTGATATTTTCCCGCGTTATGTCT	251	
QY	263	gtacacaacttctaatgaagaagcttgcagatgttatggctgtagtgcacaaagctagaag	322	
Db	252	GTAACACATCTTATTAAGATTAAGATTTCAATGTATTTAGGGCTGATGCCAAAGAGCTGAGAG	311	
QY	323	aa-taaagacaaatagtgtgaagacaaatataaa-gtttctaaacttcataccggagag	380	
Db	312	AACATAAGCAATATATGTGGAGGCCAATATTAACGTTTCTATCACTTCATACCGGGAG	371	
QY	381	atggagcatttgaagacagctccgcgaagaagctcacccaagtgcttcttcctagttatatt	440	
Db	372	ATGAGACATTTGAGACAGATTCGCCAAGACGTCACCAAGTGTCTTTCTTAAGTATTTAT	431	
QY	441	tcaatccacctttgccaactactggtctcttgcctaaagtacccgtttcccaagcaa	500	
Db	432	TCGATTCACACTTTTCCCAACATCGGTGTCTTTGCTAATGTACCTGTTTCCAGCGAA	491	
QY	501	ctactgacagcagcattcttcggaccccaaaacaacaactgtattcttagatactatcat	560	
Db	492	CTACTGATACGCGATTTCTTGAGACCCCAAAACACAACTGATTTCTTAGATATCTATCAT	551	
QY	561	gcttcccggaagcagctccaccaccagaaa-ttattagttatttagaaaagtlcatccctct	619	
Db	552	GCATTCCGAGACAGTCCACACCCAGAAACTTATTTAGTATTTAGAAAGGTCAATCCCTCT	611	
QY	620	catttctgtagcagagctccggtggtgcgtctgcacagatctgtgcaccaagataaacgctg	679	
Db	612	CATTTCGATGACGGAGACTCCGGTGGCGTCTGCACAAATCTTGACACCAAGATATACAGGTGG	671	
QY	680	taaccaccacagaaatcatgatatcttgctcgtctggagagatggttctctaacatccctct	739	
Db	672	AACCCACCCAGCAATATGATATCTTGGCTCTGAGAGAGTG-TACTCTAATCATCTCTCT	730	
QY	740	gggcgatgaacaaactccagagcttgcacgttgaagccttgagccgggccaatgtcttccac	799	
Db	731	GGCGATGAACCAACTCCAGGCTTTGAAGTGTAAGTCTTTGAGCCGGSCCATGCTTCTCAC	790	
QY	800	atcttactctccctccctctgttagagcatcgtttagaag-actcatalacaactgtagtc	858	
Db	791	ATCTTACTCTCCCTCCCTCTGCTTGTGAGGACATTCGTTGAAGCATCTTTATACACTGGGATCA	850	
QY	859	accacactcgac 869		
Db	851	CCCAACTGAC 861		
RESULT 6				
LOCUS	BG819799	849 bp	mRNA	EST 22-MAY-2001
DEFINITION	602878184.F1 NC1 CGAP_Brn67 Homo sapiens cDNA clone IMAGE:4932538			
ACCESSION	BG819799			
VERSION	BG819799.1	GI:14167386		

KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE	1 (bases 1 to 849)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/.
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
	Contact: Robert Strausberg, Ph.D. Email: cgaubs@remail.nih.gov Tissue Procurement: David N. Louis, M.D. cDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLN at: http://image.lnl.gov plate: LLM10858 row: h column: 11 High quality sequence stop: 820.
FEATURES	Location/Qualifiers
Source	1..849
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="IMAGE:4932538"
	/clone_lib="NCI CGAP Brn67"
	/tissue_type="aneurysmal oligodendroglioma with lp/19q loss"
	/lab_host="DH10B (T1 phage-resistant)"
	/note="(Organ: brain; Vector: pCMV-SPORT6; site_1: NotI; site_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2.3 kb. Constructed by Life Technologies. Note: this is a NCI CGAP library.)"
BASE COUNT	205 a 216 c 191 g 236 t 1 others
ORIGIN	
Query Match	34.9%; Score 738.2; DB 11; Length 849;
Best Local Similarity	97.0%; Pred. No. 7.8e-118;
Matches 815; Conservative	0; Mismatches 19; Indels 6; Gaps 6;
Db	1 ctgtgaagatggcgctctccaagggtgctgctggctgctgctgctgctgctgctgctgcaag 60
	5 CTGTGAAGATGGCGCTCCCA-GGATGCTGGCGTCGCGCTGNGTGGGCTCGCAG 63
OY	61 tcacccctggacaattttgtcacccggaggctgcacttgctgccttcgctcgcttgg 120
Db	64 TCACCCCTGGACA-TTTGTACCCCGAGAGCTCAACTTGCTGCTCGCTGGCTGGG 122
OY	121 ggagccctggctctcaaaagttaacctcttcaccaaggcagaatgtgaagaactgatgt 180
Db	123 GGAGCCCTGGCTCTCAAAAGCTTCAACCTTCTGCCAAGGCAATGTGAAGAAGCTGATGT 182
OY	181 ctatatgtgtaaccagaacaagaacgatataatggagaatataccacgltctcttggtcgtc 240
Db	183 CTTATGTGGTAACCAAGAACAAAGCATTAAGGGAATAACCATGCTGTTCTTGGGTCGTC 242
OY	241 attcccccgctcttatatacctgtacacaaaattctaagaagaattcgagatgttatgg 300
Db	243 ATTTCOCOCGCTTATGTCTGTACACATATTCTCAAGAAAGATTGCAAGATGTATGGG 302
OY	301 ctgatgccaaaaagctcagaagaataaagaacaatatatgtggaagacaacatatagaatttc 360
Db	303 CTGATGCCAAAAAGGCTGAGAGAATTAAGAACAAATATGTGGAAGCACATATATAAGTTTC 362
OY	361 atcaacttccalaccggagatgtagcatltyagacagttccggccaagaagtcaccaagt 420
Db	363 ATCAACTTCCATACCGGAGATGGAGCATTTTAGACAGTTTCCGCCAAGACGTACCACAAAT 422
OY	421 gcccttccctaatgatattatccattcacaccttttgcacaacaccggtgctctcttgcaa 480
Db	423 GTCCTTCTTAGCTATATTTTCATTCCACCTTTTGGCAACTACTGCTGCTTCTTGGTAA 482

OY	1181	ccttctaagtcgtgtaggaaggcagaagaaagcaagggccatgcgtcttaacagatgagcac	1240
Db	61	CTGTAAAGTGTGTTGGGAAGGACAGAAGAGAGCAGGGGCCATGGCTTCACACCAATGGCAC	120
OY	1241	acctctgaggaaactcgacagacaattccctccaagctagaacttgaaacaacaaccctctgttag	1300
Db	121	ACCTGTGGGAAGTGCAGACATTCCCTCTCACAGCTAAGAACTGAAACAACACCCTTGCTAG	180
OY	1301	gggtttgtccgttgtgagagtcatcctgtccccataaatactaatagctgtgaaactgagc	1360
Db	181	GGGTGGTCCGTGTGAGAGTGTCAATCCGTCCCCCATTAATATCTAATATGCTGGAAGTGGAC	240
OY	1361	agagagcttaactgagggcctttaactgtagtgtttcaagtctatgatctctggaagtaagctt	1420
Db	241	AGAGAGCTTACTAGGGCTTTTACTGTGANTGTTCAGTTCAATGTCCTTAGGAAGTACGCTT	300
OY	1421	ttagcccaagttggaalcccattaattggtcttaagagctagatccaactccalcaltgtaactacat	1480
Db	301	TTCGCCCAAGTGGGAATCCTTATTTGGCTTTAGAGCATGATCCACTTCAMTGTACTTAAT	360
OY	1481	ctgtgaggttttgcgtgtgtcgtcttgaaaatttttgctgtgtgaaacaagcaactcccttg	1540
Db	361	CTGTGGGTTTTGTGTGTGCTCTTGTGAAAAA-TTTTGGCTGTGGAAAACAGCACTCCTTTTG	419
OY	1541	gcttgagcaactgtgtlccaatgatctactgttggtgttlcccaccaactcttccttgatatga	1600
Db	420	GCTGGAGCACTGTGTGCATCGATGATGACTGGGTGTTCCTCCATCCTCTTGTGATATGA	479
OY	1601	ccaaaaataactgctgttltgtlltlltgcacctictaactgagcaatggctaacaactctt	1660
Db	480	CCAAAAATCAAGTTG-TATGTATATTGTACACTTCACATGGCATGGGGCTAACACACTTCTTT	538
OY	1661	tccaagac-ccttgaacaccttttctgaatgagtgtaactgtgaagaaatcttatgtgaaa	1719
Db	539	TTCMAACCCCTTGACACACCTTTTCTGTGATGGATGAACTGACGAATAATTCTATCGGAAA	598
OY	1720	agataacagaaagtaacaaagtgccttcttgaacccctcccaatgcttctagaagcttaact	1779
Db	599	AGATAACAGGAAGTACAAAGTGGCTCTTGACCCCTTCCTCAATAGTTTCTAGGCTTACACTCT	658
OY	1780	ccatgtgccttctcgtggcctgtatataacagccctctgtgagatcctcaactctgctgccca	1839
Db	659	CCATTTGCTTTTCTGGCGGTATTTACACCCCTCTGTGGATCTTCAACTCTGTGCCCTCCA	718
OY	1840	ctgtatcagcagcagcccaactgtaac-tgacagtggtgcctctctctgagccatgagatca	1898
Db	719	CTGTATGTCAGCAGATGCCAATCTTAACTTGACAGATGCGCTTCTCTGGSGCATGGATCCA	778
OY	1899	caactctaaagtactaataactgagc	1923
Db	779	CACCTGTCAAGTCTAATTAATCTGGC	803
RESULT	10		
BIO91204			
LOCUS	BIO91204	850 bp mRNA	EST
DEFINITION	602856034P1 NIH_MGC_10 Homo sapiens cDNA clone IMAGE:4597313 5'		20-JUN-2001
ACCESSION	BIO91204	mRNA sequence.	
VERSION	BIO91204.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	NIH-MGC http://mgc.nci.nih.gov/		
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)		
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rgapds-remail.nih.gov Tissue Procurement: ATCC		
	5		

FEATURES									
source	1.	850	Location/Qualifiers	row: c	column: 10	stop: 788.	High quality		
BASE COUNT	191 a	208 c	213 g	238 t					
ORIGIN									
Query Match	33.7%	Score 713.4	DB 11	Length 850					
Best Local Similarity	95.1%	Pred. No. 1.5e-113							
Matches 813	Conservative	0	Mismatches 31	Indels 11	Gaps				
QY	910	aggaaatgaatcggcttcttatacccgggcgctgaattccagatattggtgaaga	969						
Db	1	AGGAAATGAATTCGGCTTTATCTCCGCGCTGAATTCAGCATATTTGGTGAAGATA	60						
QY	970	ggatgcgaacttgcctggagaaatgagctgcaattctcctgcagccttgaagaagctgagc	1029						
Db	61	GGTGTGCAACTTGGCGTGGAGAGATGGCTTCAGATTTCCTGCAGCCGTGAAGAAGCTGAGC	120						
QY	1030	tgtctctctgtctgaacaacgctggtcctgctctccaccacacttaccttgggaagagcgct	1089						
Db	121	TGTCTCTCTGTCTGCGCAACAGTGTCCTGCTCTCCACCAACTACCTTGGGACAAGCGCGT	180						
QY	1090	gaattgaaccatlgagagcggatgagcatgtccctcagtcgatatgatatagcagtcgaggaac	1149						
Db	181	GAATGACCATGAGGAGCGGATGGCATTTGTCCTCAGCGTATATGATATAGCAGTCGACGAAC	240						
QY	1150	aaacagcacttgcacgcaaaagctctgtgtgtactgttaagtgtgtggagagca-gagagag	1208						
Db	241	AAACAGCACTTCCGACGAAAGTCTGTGTACTGTAAAGTGTGTGGAGGCGCACGAGAGAG	300						
QY	1209	gagcagggggccatlggctctcacagatlggcaaccttgygggaactgcgcagaattcctctc	1268						
Db	301	GAGCAGGGGCCATVGGGCTTCACAGCATGGCACATCTGTGGGAACATGCAAGACATTCCTCTC	360						
QY	1269	acagatagaactgaacaaacccctcttgccttaagggtgtgtcgtgtgaggtgtcatcctgt	1328						
Db	361	ACAGCTAGACATGGAACAACACCCCTTTGCTAGAGGGTGGTCCGTGTGAGGTGATCACTCTGT	420						
QY	1329	ccccctcataattacttaataagcttgaacttgcagcagcctcctactgagcttcttactgtga	1388						
Db	421	CCCCCTCATTAATTAATAATAGCTGGAACCTGGCAGCAGCCTCACTACGGGCTTTTAATGTGA	480						
QY	1389	tgtgtcagctcatctccttagaagtagcctttggccccaggtgggaattccttatttggc	1448						
Db	481	TGTGTTCAGTTCAATCTCTAGGAATCTAGCTTTTGGCCCCAGGTGGGAATCCCTTAATTGGC	540						
QY	1449	ttagagcatcatccattcatatgttacttacaacatctgttggtttttgttgtctgttagaa	1508						
Db	541	TTAGAGCATGATCCATCTTCATAGTTACTACATCTGTGGGTTTT--GTGCTGCTGTAGAA	598						
QY	1509	aatttttgcctgtgtgaaacagcactcctcttgcctgtgagcaactgtgtctcatgtcatgtac	1568						
Db	599	AA-TTTCGGCTGGTGAAGAAACAGCACTCCTTTGGCTGTGGAGCACTGTGTGCATGTGATGAC	657						
QY	1569	tttagatgtcttcacacatcctttcttgatataatgcaaaaaataaattttttgtttttttgt	1628						

Query Match	33.6%	Score 711.8;	DB 11;	Length 844;
Best Local Similarity	96.5%;	Pred. No. 2.8e-113;		
Matches 802;	Conservative 0;	Mismatches 22;	Indels 7;	Gaps
QY	I	cftgtgaagatgagcgcctcccaagtgtgcbtgytgctgctgacgtgtggagctgcgaag	60	
Db	10	CTGTGAAGATGGCGGCGCCCA-GGTGTCTCTGCCTCGCTCGCTGTTGTGGGCTCGCAG	68	
QY	61	tcaacccttgaaattttgtcacaccggaagctgcaacttggcttcctcttgacctgttgg	120	
Db	69	TCAACCTTGACA-TTGTTCACCCGGAAGCTGCAC-TGGTGCCTCTGGCTGTGGCTGGG	126	

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 967)	Mammalia; Eutheria; Primates; Catarrhini; Homiinae; Homo.	NIH-MGC http://mgc.nci.nih.gov/ .	National Institutes of Health, Mammalian Gene Collection (MGC)	
	Unpublished (1999)	Contact: Robert Straussberg, Ph.D. Email: c9q9bs-r@mai.nih.gov		
	Tissue Procurement: ATCC	CDNA Library Preparation: Life Technologies, Inc.		
	CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be			


```

Db 301 CTTGGCTGAATTCACGATA-TGGTGAAGATAGTGTGCAACTTGGCTGGAGCAATGG 359
Qy 996 ctgcagattctcgcagcctgaagaacgtgctgtctctctctctcgcacaaagtgtc 1055
Db 360 CTGCAGATTCTCTGCAGCCTGAAGAAGACCTGAGCTGTCTCTCTGCTTCACAACTGTGTC 419
Qy 1056 ctgctctccacaactaccttggagacaagcgctgaatgaacatgagcgagatgacat 1115
Db 420 CTGCTCTCCACCAACTACTTGGGACAGGGCGCTGAATGACACATGAGCGGATGGCAT 479
Qy 1116 gtccgcaggtgtatagatagcagtcaggaacaacagcacttgcagcaaaagtctgt 1175
Db 480 GTCTCGAGTGTATGATATAGCACTGCAGACAAACAGCACTTGGCCAGCAAGTCTGT 539
Qy 1176 gtctactgtatagtggtggagcaagagagagagagagagagagagagagagagagat 1235
Db 540 GTGACTGTTAAGTGTGTGGGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 599
Qy 1236 ggcacacctgtggaaactgcagacatctctcacaagctgaactgaacaacacctctt 1295
Db 600 GGCACACATGTGGGAGACGTGACATTCCTCTCAGAGTGAACCTGAACAAACCTCTT 659
Qy 1296 gctaggggtgtgcgtgtgagtggtctcctgtccctcata 1338
Db 660 GCTAGGGGTGTGCTGTGGAGGTGTCTCTCTCCCTCATATA 702

RESULT 14
LOCUS BG473855 714 bp mRNA EST 21-MAR-2001
DEFINITION 602515590F1 NIH_MGC_16 Homo sapiens cDNA clone IMAGE:4647143 5',
VERSION BG473855 mRNA sequence.
KEYWORDS BG473855.1 GI:13406132
SOURCE EST.
ORGANISM human.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 714)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLES National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999).
COMMENT Contact: Robert Strausberg, Ph.D.
Email: sgabds@email.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
Plate: L1CM1423 row: d column: 24
High quality sequence stop: 713.
Location/Qualifiers
1. 714
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4647143"
/clone_11b="NIH_MGC_16"
/tissue_type="retinoblastoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: eye; Vector: pOT7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5'
adaptor: GGCAGAG(G). Library constructed by Ling Hong
in the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-CDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."
BASE COUNT 165 a 166 c 191 g 192 t
ORIGIN

```

```

Query Match 32.4%; Score 686; DB 11; Length 714;
Best Local Similarity 98.5%; Pred. No. 8.2e-109;
Matches 703; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

Qy 890 gattggcagctgactgtctcaggaagtaaatcgctgttattctcgtgtgactgaatc 949
Db 1 GATTGGCAGCTGACTGCTCAGGAAGTAATAATGGCGTTGTTACTCTCCGTGGCTGAATTC 60
Qy 950 taacgatattgtgaaatagtggtgtcgaacttggcttggaggaatgtgtgagattctcgt 1009
Db 61 TAGCATATTGTGTAAGATATAGTGTGCAACTTGGCTGGGAATGGCTGAGATTTCTGTG 120
Qy 1010 cagctgaaagaagctgtagctgtctctctgtctgtcacaagtggtlctcgtctccacca 1069
Db 121 CAGCTGAAAGAAGCTGAGCTGTCTCTCTGTGTCGACAAAGTGCTCTGCTGCCACCA 180
Qy 1070 ctaccttggagacaagcgctgaaatgaacaaatggaagcggatggcattgtctcagtcgt 1129
Db 181 CTACCTTTGGACAAAGCGCTGAATGAACCATYGAAGGAGGATGGCATTTGCTGCAGTCGA 240
Qy 1130 taatatagcagtgcaagaaacaacagcacttgcagcaaaagtctgtgttactgttaagt 1189
Db 241 TAGTATAGCAGTGCAGAGCAACAAACACACTTGGCCACAAAGTGTGTACTGTTAAGT 300
Qy 1190 gttgtgagagcagaagagagagagagagagagagagagagagagagagagagagag 1249
Db 301 GTGTGGAGGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 360
Qy 1250 aactgcagacatctctctacagctgaactgaactgaactgaactgaactgaactgaact 1309
Db 361 AACTGCAGACATTCCTCTCTACAGCTGAACCTGAACAAACCTCTTGTGAGGGGTGTC 420
Qy 1310 gttgtgagtgcaatctctgtccctcctaataactaataactgtcgtgagagcagcctc 1369
Db 421 GTGTGAGGTGTATCTCTGTCCCTCATTAATTAATAATAGTGAAGAGGACAGAGCTC 480
Qy 1370 tactggcttcttactggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1429
Db 481 TACTGGGCTTTTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 540
Qy 1430 gtgggaatccttacttggcttgaactgaactgaactgaactgaactgaactgaactgtgt 1489
Db 541 GTGGGAATCCTTATTTGGCTTAGGACTGATCCAAATTCATGTTACTTACATCTGGGTC 600
Qy 1490 ttgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1548
Db 601 ACAGTAGTAGCTGTAGAAAATTTAGGCTGTGTGAAGACAGACACTCTTGGCTGGCAGC 660
Qy 1549 acttgttcatcatgactgtactgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1602
Db 661 ACTTGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 714

RESULT 15
AL557670 702 bp mRNA EST 16-FEB-2001
LOCUS AL557670 LTI_FL012_Tc1 Homo sapiens cDNA clone CSDB007Y013 5 prime
DEFINITION , mRNA sequence.
ACCESSION AL557670
VERSION AL557670.1 GI:12901503
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 702)
AUTHORS Li, W.-B., Gruber, C., Jessee, J., and Polayes, D.
TITLES Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 20, 2001, 17:25:04 ; Search time 99.76 Seconds

(without alignments)
4808.344 Million cell updates/sec

Title: US-09-868-474-1

Perfect score: 2118

Sequence: 1 ctgtgaagatgctgctcc.....aaaaaaaaaaaaaaaaaaaaa 2118

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 08

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, NA:*

- 1: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
- 2: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
- 3: /cgn2_6/ptodata/2/1na/6A.COMB.seq:*
- 4: /cgn2_6/ptodata/2/1na/6B.COMB.seq:*
- 5: /cgn2_6/ptodata/2/1na/PCUTUS.COMB.seq:*
- 6: /cgn2_6/ptodata/2/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55	2.6	7218	1 US-08-232-463-14	Sequence 14, Appl
2	54.6	2.6	3138	1 US-07-867-106-4	Sequence 4, Appl
3	51.2	2.4	1582	3 US-08-545-196B-10	Sequence 10, Appl
4	51.2	2.4	1582	3 US-08-545-196B-12	Sequence 12, Appl
5	51	2.4	2671	6 5168051-9	Patent No. 5168051
6	50.2	2.4	2744	3 US-09-071-101-1	Sequence 1, Appl
7	50.2	2.4	2744	3 US-09-369-618-1	Sequence 1, Appl
8	50.2	2.4	2744	3 US-09-369-617-1	Sequence 1, Appl
9	49.8	2.4	1522	4 US-09-413-574-1	Sequence 1, Appl
10	49.8	2.4	3736	2 US-08-480-473B-1	Sequence 1, Appl
11	49.8	2.4	3736	3 US-08-915-213-1	Sequence 1, Appl
12	49.8	2.4	3736	3 US-09-148-547-1	Sequence 1, Appl
13	49.8	2.4	3736	4 US-09-235-217-1	Sequence 1, Appl
14	49.8	2.4	3736	5 PCT-US96-10251-1	Sequence 1, Appl
15	49.6	2.3	2296	4 US-08-496-841C-137	Sequence 1, Appl
16	49.4	2.3	375	3 US-08-946-026-23	Sequence 13, App
17	49.4	2.3	2567	3 US-08-993-260-4	Sequence 23, Appl
18	49.4	2.3	2628	1 US-08-143-219-1	Sequence 1, Appl
19	49.4	2.3	3581	2 US-08-738-349-1	Sequence 1, Appl
20	49.2	2.3	1540	3 US-08-977-001-2	Sequence 2, Appl
21	49.2	2.3	2852	3 US-09-027-137-2	Sequence 2, Appl
22	49	2.3	3527	2 US-08-909-965C-7	Sequence 7, Appl
23	49	2.3	3637	1 US-08-445-640-3	Sequence 3, Appl
24	49	2.3	3637	1 US-08-170-558-3	Sequence 3, Appl
25	49	2.3	3637	3 US-08-447-314-3	Sequence 3, Appl
26	49	2.3	3637	3 US-08-445-611-3	Sequence 3, Appl
27	48.8	2.3	966	1 US-08-514-014-7	Sequence 7, Appl

ALIGNMENTS

28	48.8	2.3	966	2 US-08-833-823-7	Sequence 7, Appl
29	48.8	2.3	1393	1 US-08-174-467-18	Sequence 18, Appl
30	48.8	2.3	1393	3 US-08-452-071-18	Sequence 18, Appl
31	48.8	2.3	1804	2 US-08-504-459-5	Sequence 5, Appl
32	48.4	2.3	1315	4 US-09-721-822A-10	Sequence 10, Appl
33	48.2	2.3	467	2 US-08-841-349-18	Sequence 18, Appl
34	48.2	2.3	1931	3 US-09-019-942-2	Sequence 2, Appl
35	48.2	2.3	2255	1 US-08-871-572B-3	Sequence 3, Appl
36	48	2.3	208	1 US-08-686-878A-37	Sequence 37, Appl
37	48	2.3	1772	2 US-08-960-022-13	Sequence 13, Appl
38	48	2.3	2301	4 US-09-232-191-8	Sequence 8, Appl
39	48	2.3	2301	4 US-09-232-200-8	Sequence 8, Appl
40	48	2.3	2301	4 US-09-232-197-8	Sequence 8, Appl
41	48	2.3	2710	4 US-09-232-200-44	Sequence 44, Appl
42	48	2.3	2710	4 US-09-232-200-70	Sequence 44, Appl
43	48	2.3	2710	4 US-09-232-197-44	Sequence 44, Appl
44	48	2.3	2710	4 US-09-232-197-70	Sequence 70, Appl
45	47.8	2.3	3785	1 US-08-485-618-98	Sequence 98, Appl

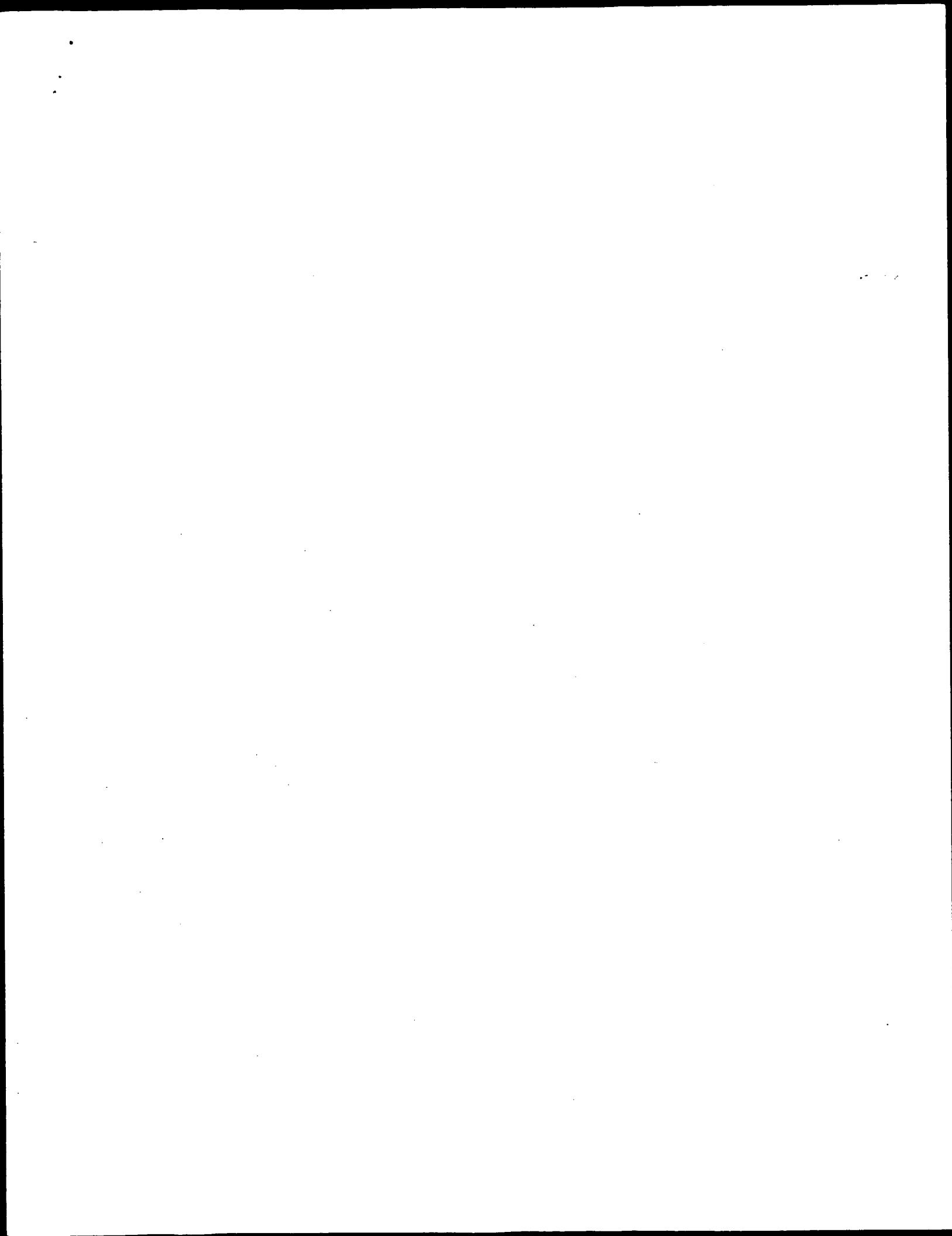
RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEFFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOVILPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)836-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZgpt-Fls
US-08-232-463-14

0
0
0
0
1

0
0
0
0
1

SEQUENCE CHARACTERISTICS:

QY 2094 aaaaaaaaaaaaaaaaaa 2118



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 20, 2001, 10:09:52 ; Search time 15.97 Seconds

(without alignments)
1717.147 Million cell updates/sec

Title: US-09-868-474-2

Perfect score: 1912
Sequence: 1 MALSRVCMARSAMGSAVTP.....LSLLHNVLSTNYLGTTR 360

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	520	27.2	100	2 T08763	hypothetical prote
2	150.5	7.9	398	2 C96680	hypothetical prote
3	143	7.5	485	2 T38255	hypothetical prote
4	140	7.3	770	2 T22944	hypothetical prote
5	140	7.3	784	2 T22939	hypothetical prote
6	136.5	7.1	755	2 T47806	hypothetical prote
7	136	7.1	753	2 S66710	probable membrane
8	131.5	6.9	497	2 T21579	hypothetical prote
9	131	6.9	481	2 T42226	hypothetical prote
10	119.5	6.2	454	2 S69017	probable membrane
11	101.5	5.3	431	1 C69689	response regulator
12	99	5.2	3795	2 T00831	hypothetical prote
13	95	5.0	369	2 B69799	hypothetical prote
14	92.5	4.8	728	2 A81385	salicylate 1-mono
15	92	4.8	656	2 H71867	probable ATP /GTP
16	91.5	4.8	1120	2 S46111	methionine--tRNA
17	90	4.7	880	2 S61908	probable membrane
18	89	4.7	276	2 S35270	DNA-binding protei
19	88.5	4.6	287	2 J01743	DNA-damage repair
20	88	4.6	1163	2 T24855	hypothetical prote
21	87	4.6	314	2 G69818	hypothetical prote
22	86.5	4.5	735	2 D96737	CMF-binding factor
23	86.5	4.5	254	2 T43494	ARE1-like protein,
24	86.5	4.5	957	2 C69463	type I restriction
25	86	4.5	1715	2 T06145	disease resistance
26	85.5	4.5	679	2 T20713	hypothetical prote
27	85.5	4.5	2231	2 S53416	SENI protein - Yea
28	84	4.4	561	2 F66561	arginyl tRNA trans
29	84	4.4	561	2 C72063	arginyl-tRNA synth

30	84	4.4	650	2 A64572	methionine--tRNA
31	84	4.4	701	2 G96737	ARE1-like protein,
32	84	4.4	2165	2 T21371	hypothetical prote
33	83.5	4.4	339	2 T15251	hypothetical prote
34	83.5	4.4	465	2 T48374	UDP-glucosyltrans
35	83.5	4.4	603	2 B86608	exonuclease ABC,
36	83.5	4.4	603	2 F72017	hypothetical prote
37	83.5	4.4	772	2 T16474	hypothetical prote
38	83	4.3	1433	2 A71444	probable LTR retro
39	82.5	4.3	2199	2 T40006	Cdc20P - fission y
40	82	4.3	248	2 H83111	hypothetical prote
41	82	4.3	1596	2 T24146	hypothetical prote
42	81.5	4.3	816	2 T25555	hypothetical prote
43	81.5	4.3	884	2 F96659	protein F2K11.26 f
44	81.5	4.3	1511	2 T42711	sulfonylurea recep
45	81.5	4.3	1546	2 T42728	sulfonylurea recep

ALIGNMENTS

RESULT 1
T08763
hypothetical protein DKFZp586A011.1 - human
C:Species: Homo sapiens (man)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 04-Mar-2000
C:Accession: T08763
R:Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wleemann, S.
submitted to the Protein Sequence Database, May 1999
A:Reference number: 216468
A:Accession: T08763
A:Molecule type: mRNA
A:Residues: 1-100 <KOE>
A:Cross-references: EMBL:AL050286
A:Experimental source: adult uterus; clone DKFZp586A011
C:Genetics:
A:Note: DKFZp586A011.1
C:Superfamily: human hypothetical protein DKFZp586A011.1

Query Match 27.2%; Score 520; DB 2; Length 100;
Best Local Similarity 100.0%; Pred. No. 3.7e-38;
Matches 100; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 261 MLTSLPPLRLRHUKHTTIVHOLDKALAKLGIGQLPADEVKSACYLRGINSTHIGED 320
DB 1 MLTSLPPLRLRHUKHTTIVHOLDKALAKLGIGQLPADEVKSACYLRGINSTHIGED 60
QY 321 RCRTWLGEWLQISCSLKEAEISLLHNVLSTNYLGTTR 360
DB 61 RCRTWLGEWLQISCSLKEAEISLLHNVLSTNYLGTTR 100

RESULT 2
C96680
hypothetical protein F5114.7 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96680
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federzpiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
Nure, N.F.; Hughes, B.; Hutzar, L.
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, T.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: C96680
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-398 <SPD>
A:Cross-references: GB:AE005173, NID:92190544, PIDN:ABB60908.1, GSPDB:GN001444
C:Genetics:
A:Gene: F5114.7
A:Map position: 1

Query Match	7.9%;	Score 150.5;	DB 2;	Length 398;
Best Local Similarity	21.2%;	Pred. No. 2.2e-05;		
Matches 54;	Conservative 55;	Mismatches 119;	Indels 27;	Gaps 7

```

QY  92 GUMAMAAKXARIRIKTMMKHNKIHQDPYREMEHLOFRODQVTKCFLGISIPFAN 151
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 144 GKLLIMADRIISVRLVLYL-----ANGKLSRRERQOQLTRTADIIRLVAVFELIVPME 199
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 152 YVFFLLMYLFFPKOLLIRHFWTPKQOQDFLDIYHAFKROSH-----PEIISTLEKVI 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 200 FLFPVFLKLFKRWML-----PSTFQDMKKKEEALKRLNARMEXAKFLDQTVKMAKEV 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 203 PLISDAGLRMRILTDL---CTKLORECTHRAIDHILALRECFSNHPHGMQOLALHVKALSR 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 QTSRSGEIKKTAEDIDGEMTKVRKRGVGSNDIEILFAKLF-NDLELTLDNIRSRKLVNNCK 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 AMLLTSYLPPLRLRHLKHTTIVYHOLDKALAKIGLITGOLTAOEKVSACYLRGLNSTHIGE 319
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 YMGISPEGDAYLRYMLRKRLQEIKKDKLLKAEVGSLSSEALEROACREGRM--LQDGS 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 320 -DRCRTWLGEMLOIS 333
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 VEEMREQLVDWMDLS 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT      3
T38255
hypothetical protein SPAC23C11.17 - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38255
R:Brown, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A:Reference number: Z21781
A:Accession: T38255
A:status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-485 <BRO>
A:Cross-references: EMBL:Z98559, PIDN:CAB1166.1, GSPDB:GN00066, SPDB:SPAC23C11.17
A:Experimental source: strain 972h-, cosmid c23C11
C:Genetics:
A:gene: SPDB:SPAC23C11.17
A;Map position: 1
A;Introns: 26/p1

```

Query Match	7.58;	Score 143;	DB 2;	Length 485;
Best Local Similarity	21.08;	Pred. No. 0.00013;		
Matches	66;	Conservative	71;	Mismatches 150; Indels 28; Gaps 11

```

Oy 40 PRSCKLHLSBP -ADVNYLNSVYVTKIANGVYHFLRHPREYLYTTFMKLOJLMA 98
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 90 PRTSKYKAPQVAEMETPTTVVKKPSI---WQVKGSG-----VLH--FMDGCKTLGV 138
Oy 99 DAKKARIKTNNMKHNIKTNOLRYEMENHLRFODVTKCFLGIISTPPANVLVLLM 158
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 139 EIKISSKL--YKMAVGY-ELTRRESROLPTFLDRLRPVSYVAVPAAELLIPAV 194
Oy 159 YLRPRLLIHRFTPAQOQDDELDIYHAFKOSHPELISYLEKVI-----PLISDAGLRMR 213
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 195 KLRP-NLBPSTEDADOK---EAKKAOLRIKTRNEVSNMNLSTLSKGFPTSNNETRESKE 249
Oy 214 LTLPLCTKIQ-RGTHPAIDHILRECFSNHPLGMMOALHNVKALSRLMLTSTYLPPL 272
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
Db 250 FROFPQKVRHSQSPERBELIEVCYKFKD-DITLDNLSPAOVLVAMCRVMYNNIAGTDPBL 308

```

[illegible]

```

RESULT 4
T22944
hypothetical protein F58G11.1b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T22944
R:Percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19640
A:Accession: T22944
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-770 <WIL>
A:Cross-references: EMBL:Z81094; PIDN:CAB03156.1; GSPDB:GN00023; CESP:F58G11.1b
A:Experimental source: clone F58G11
C:Genetics:
A:Gene: CESP:F58G11.1b
A:Map position: 5
A:Introns: 44/2,184/3; 360/2; 708/3; 743/3

```

Query Match	7.38;	Score 140;	DB 2;	Length 770;
Best Local Similarity	23.28;	Pred. No. 0.0004;		
Matches	57;	Conservative	52;	Mismatches 113;
				Indels 24;
				Gaps 73;

[illegible]

```

RESULT      5
T22939
hypothetical protein F58G11.1a - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C/Accession: T22939
R/Percy, C.
submitted to the EMBL Data Library, October 1996
A/Reference number: Z19640
A/Accession: T22939
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-784 <NTL>
A/Cross-references: EMBL:Z81094; PIDN:CA803150.1; GSPDB:GN00023; CESP:F58G11.1a
A/Experimental source: clone F58G11
C/genetics
A/Gene: CESP:F58G11.1a
A/Map position: 5

```

A:introns: 44/2; 184/3; 360/2; 708/3; 757/3

Query Match

Best Local Similarity 7.3%; Score 140; DB 2; Length 784;
Matches 57; Conservative 52; Mismatches 113; Indels 24; Gaps 7;

QY 120 LPYREMEHLROFRODVTKCLFLGIISIPFANNYLVLFLMLTFRQLLRHFKTHHTVYHQL 179
DB 129 LSRERQOLVATVSDFLVLPFSEFIIIVPMEALPFIPLFGMLPSPFOESSKEE--- 185
QY 180 LDYHAFPKQ--SHPEIISYLEKVIPLI-----SDAGLRWRLTDLCTYIO-RGTH 226
DB 186 ----EKRRKQVKLRVEMAKFLDTEIEIGLERKTRNKSESTRSLFAL--FIKVRNEGCV 239
QY 227 PAINDILALRECFSNHPLGKNOLOALHVKALSRAMLLTSTLPPLLRHKLKTHHTVYHQL 286
DB 240 VSNELLKFKSLPEDE-ITLDNLSKQOLRSLCKLMSINSLSPEILLRQOLNMKIRELKAD 298
QY 287 DKAL-AKLGIGQLAOEVKSACTLRGLNSTHIGEDRCRTWLGEMLOISCKEAEISLL 345
DB 299 DKQIAEGVDALSSIDLQSAACRARGMRAIGVSEERLKQLVQWLELSLNDKVPALLL 358
QY 346 HNVVLL 351
DB 359 SRTLYL 364

RESULT 6

T47806

Hypothetical protein F24G16.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T47806

R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Lemcke, K.; submitted to the Protein Sequence Database, February 2000

A:Reference number: 224477

A:Accession: T47806

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-755 <DAN>

A:Cross-references: EMBL:AL138647

A:Experimental source: cultivar Columbia; BAC clone F24G16

C:Genetics:

A:Map position: 3

A:introns: 194/2; 298/3; 347/3; 377/2; 408/2; 450/3; 472/2; 530/3; 595/3; 611/3; 660/3; A:note: F24G16.90

Query Match

Best Local Similarity 7.1%; Score 136.5; DB 2; Length 755;
Matches 50; Conservative 56; Mismatches 123; Indels 25; Gaps 6;

QY 92 GLQMLMADAKKARRITNMKNHNIKPHOLPYREMEHLROFRODVTKCLFLGIISIPFAN 151
DB 218 GTKLMDADTRISSRLILK-----AGKSLSRERQOLTRTADIFRLVFAVFIIPFWE 273
QY 152 YIVVFLMLTFRQLLRHFKTHHTVYHQLKALAKLIGQLAOEVKSACTLRGLNSTHIG 209
DB 274 FLVPLFKLFPNML-----PSTFODKKMEFEALRKLLAREVAKFLQETARAKMAEV 326
QY 210 LMR-----LTDLCTKIORGTHPAIHDLALRECFSNHPLGKNOLOALHVKALS 259
DB 327 KHSRTGEVQOTADDEFLDKVRKGOIVNDELGLPAKLF-NDELTLTNISRPVLSMCK 385
QY 260 AMLLTSTYLPPLLRHKLKTHHTVYHQLKALAKLIGQLAOEVKSACTLRGLNSTHIG 319
DB 386 YMGISPYGDAVLRWLRKRLRSIKEDDKLIRAEVDSLSAELELRDREGMGL-VSV 444
QY 320 DRCRTWLGEMLOIS 333
DB 445 EEMRQOLRDMMDLS 458

RESULT 7

S66710

probable membrane protein YOL027c - Yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O2165

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 05-Nov-1999

C:Accession: S66710

R:Habbib, B.; Hattenhorst, U.; Hollenberg, C.P.; Ramezani Rad, M. submitted to the Protein Sequence Database, July 1996

A:Reference number: S66703

A:Accession: S66710

A:Molecule type: DNA

A:Residues: 1-573 <HAB>

A:Cross-references: EMBL:Z74769; NID:g1419812; PID:e251846; PID:g1419813; GSPDB:GN000

A:Experimental source: Strain S288C

C:Genetics:

A:Gene: MIPS:YOL027c

A:Map position: 15L

C:Keywords: transmembrane protein

F:139-155/Domain: transmembrane #status predicted <TM>

Query Match 7.1%; Score 136; DB 2; Length 573;
Best Local Similarity 23.2%; Pred. No. 0.00062;
Matches 68; Conservative 58; Mismatches 117; Indels 50; Gaps 12;

QY 89 FMKGLMADAKKARRITNMKNHNIK---HOLPYREMEHLROFRODVTKCLFLGIIS 145
DB 94 YANGTKLLGEYELKIVSKTL-----IKFAQCYELSRERQOLRTMGDVRLRIFSAFL 146
QY 146 IPPFANYLVFLMLYLP-----ROLLRHFWTPKQOTDPLDIYHARRQ 189
DB 147 IIPFALFLPFAKLFPNLPSTYESGMDQAKRKLIE---IKKTSSEFL---HETLEE 200
QY 190 SHPEIISYLEKVIPLISDAGLRWRLTDLCTYIORGTHPAI---HD-ILALRECFSNHPL 244
DB 201 SN-LTY-----NTIENAEKKOKFLNFFRLYSAGEKGIWTFQHDIEISAIQMFKNDSV 253
QY 245 GNOLOALHVKALSRAMLTSTYLPPLLRHKLKTHHTVYHQLKALAKLIGQLAOEVK 304
DB 254 -LDNLSRQOLAMSKFMSLRFGNDNMIRQIRKLDIMNDYDIDYEGVESLSQEEY 312
QY 305 SACYRGKINSTHIGEDRCRTWLGEMLOISCKEAEISLLHNVVLLSTNYLG 357
DB 313 QACYSRKKAIGVSKEDLVNLLKWMLEL--RLRQKIPSVLM--VLSPPTFG 360

RESULT 8

T21579

hypothetical protein F30F8.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000

C:Accession: T21579

R:Milkinson, J. submitted to the EMBL Data Library, October 1996

A:Reference number: Z19443

A:Accession: T21579

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-497 <MIT>

A:Cross-references: EMBL:Z81073; PIDN:CA803034.1; GSPDB:GN00019; CESP:F30F8.3

A:Experimental source: clone F30F8

C:Genetics:

A:Gene: CESP:F30F8.3

A:Map position: 1

A:introns: 36/3; 109/1; 189/1; 235/2; 276/1; 314/3; 342/2; 384/3; 404/3; 451/3; 479/1

C:superfamily: Caenorhabditis elegans hypothetical protein F30F8.3

Query Match 6.9%; Score 131.5; DB 2; Length 497;
Best Local Similarity 21.0%; Pred. No. 0.0013;
Matches 62; Conservative 51; Mismatches 127; Indels 55; Gaps 8;


```

A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-3795 <STO>
A:Cross-References: GB:AE002093; NID:g6598526; PDB:1AAD2011.4.2; GSD:BG00139
C:Gene:
A:Gene: At2g17930; T13L16.5
A:Map position: 2
A:Map position: 201; 25873; 28772; 32373; 35573; 37673; 39673; 41973; 44171; 45873; 46773; 48773; 50773; 52773; 54773; 56773; 58773; 60773; 62773; 64773; 66773; 68773; 70773; 72773; 74773; 76773; 78773; 80773; 82773; 84773; 86773; 88773; 90773; 92773; 94773; 96773; 98773; 100773; 102773; 104773; 106773; 108773; 110773; 112773; 114773; 116773; 118773; 120773; 122773; 124773; 126773; 128773; 130773; 132773; 134773; 136773; 138773; 140773; 142773; 144773; 146773; 148773; 150773; 152773; 154773; 156773; 158773; 160773; 162773; 164773; 166773; 168773; 170773; 172773; 174773; 176773; 178773; 180773; 182773; 184773; 186773; 188773; 190773; 192773; 194773; 196773; 198773; 200773; 202773; 204773; 206773; 208773; 210773; 212773; 214773; 216773; 218773; 220773; 222773; 224773; 226773; 228773; 230773; 232773; 234773; 236773; 238773; 240773; 242773; 244773; 246773; 248773; 250773; 252773; 254773; 256773; 258773; 260773; 262773; 264773; 266773; 268773; 270773; 272773; 274773; 276773; 278773; 280773; 282773; 284773; 286773; 288773; 290773; 292773; 294773; 296773; 298773; 300773; 302773; 304773; 306773; 308773; 310773; 312773; 314773; 316773; 318773; 320773; 322773; 324773; 326773; 328773; 330773; 332773; 334773; 336773; 338773; 340773; 342773; 344773; 346773; 348773; 350773; 352773; 354773; 356773; 358773; 360773; 362773; 364773; 366773; 368773; 370773; 372773; 374773; 376773; 378773; 380773; 382773; 384773; 386773; 388773; 390773; 392773; 394773; 396773; 398773; 400773; 402773; 404773; 406773; 408773; 410773; 412773; 414773; 416773; 418773; 420773; 422773; 424773; 426773; 428773; 430773; 432773; 434773; 436773; 438773; 440773; 442773; 444773; 446773; 448773; 450773; 452773; 454773; 456773; 458773; 460773; 462773; 464773; 466773; 468773; 470773; 472773; 474773; 476773; 478773; 480773; 482773; 484773; 486773; 488773; 490773; 492773; 494773; 496773; 498773; 500773; 502773; 504773; 506773; 508773; 510773; 512773; 514773; 516773; 518773; 520773; 522773; 524773; 526773; 528773; 530773; 532773; 534773; 536773; 538773; 540773; 542773; 544773; 546773; 548773; 550773; 552773; 554773; 556773; 558773; 560773; 562773; 564773; 566773; 568773; 570773; 572773; 574773; 576773; 578773; 580773; 582773; 584773; 586773; 588773; 590773; 592773; 594773; 596773; 598773; 600773; 602773; 604773; 606773; 608773; 610773; 612773; 614773; 616773; 618773; 620773; 622773; 624773; 626773; 628773; 630773; 632773; 634773; 636773; 638773; 640773; 642773; 644773; 646773; 648773; 650773; 652773; 654773; 656773; 658773; 660773; 662773; 664773; 666773; 668773; 670773; 672773; 674773; 676773; 678773; 680773; 682773; 684773; 686773; 688773; 690773; 692773; 694773; 696773; 698773; 700773; 702773; 704773; 706773; 708773; 710773; 712773; 714773; 716773; 718773; 720773; 722773; 724773; 726773; 728773; 730773; 732773; 734773; 736773; 738773; 740773; 742773; 744773; 746773; 748773; 750773; 752773; 754773; 756773; 758773; 760773; 762773; 764773; 766773; 768773; 770773; 772773; 774773; 776773; 778773; 780773; 782773; 784773; 786773; 788773; 790773; 792773; 794773; 796773; 798773; 800773; 802773; 804773; 806773; 808773; 810773; 812773; 814773; 816773; 818773; 820773; 822773; 824773; 826773; 828773; 830773; 832773; 834773; 836773; 838773; 840773; 842773; 844773; 846773; 848773; 850773; 852773; 854773; 856773; 858773; 860773; 862773; 864773; 866773; 868773; 870773; 872773; 874773; 876773; 878773; 880773; 882773; 884773; 886773; 888773; 890773; 892773; 894773; 896773; 898773; 900773; 902773; 904773; 906773; 908773; 910773; 912773; 914773; 916773; 918773; 920773; 922773; 924773; 926773; 928773; 930773; 932773; 934773; 936773; 938773; 940773; 942773; 944773; 946773; 948773; 950773; 952773; 954773; 956773; 958773; 960773; 962773; 964773; 966773; 968773; 970773; 972773; 974773; 976773; 978773; 980773; 982773; 984773; 986773; 988773; 990773; 992773; 994773; 996773; 998773; 1000773; 1002773; 1004773; 1006773; 1008773; 1010773; 1012773; 1014773; 1016773; 1018773; 1020773; 1022773; 1
```

```

202 -----IPLISDAGLRWRI/TDI/CTKTOGCTHRAINDYAIPEGESNHHYICMNOGA 264
217 VEWL-----QPCSTDYW-DEEDTFEDRVKKEILSGFGLDEFTKHKMSQHPVIFNKLEOV 269

```

Db 270 AVEPEWKGAVIIGDAAHAGAPF-----LAOGAAMAEIAIYLAELQNHADHETALQA 324
 QY 252 LHVKALSLAMLLTGYLPPLRLHRLKTHHTVYIHQDLKALAKLGIGOLTAQEVKSACY 308
 Db 325 YKRRAPRALKVON-LSSEIVRRRLK-----GEGAEELIGEY 362

RESULT 14

A81385
 Probable ATP /GTP binding protein Cj0411 [Imported] - Campylobacter jejuni (strain NCTC C:Species: Campylobacter jejuni
 C:Date: 31-Mar-2000 #sequence.revision 31-Mar-2000 #text_change 31-Mar-2000
 C:Accession: A81385
 R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; O'Neill, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barré-Naulte 403, 665-668, 2000
 A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyf
 A:Reference number: A81250; MUID:20150912
 A:Accession: A81385
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-728 <PAR>
 A:Cross-references: GB:AL139075; GB:AL11168; NID:96967817; PIDN:CAB74247.1; PID:9696788
 A:Experimental source: serotype O2, strain NCTC 11168
 C:Genetics:
 A:Gene: Cj0411

Query Match 4.8%; Score 92.5; DB 2; Length 728;
 Best Local Similarity 22.1%; Pred. No. 5;

Matches 79; Conservative 47; Mismatches 115; Indels 117; Gaps 17;

QY 82 FYLYITFMKGLQMLADAKKARIKTMMKHNKTFHQLPYREMEHLRQFRODYTKCL-- 139
 Db 128 FKELNTINLE-LQKLSFD-----ENLKSRLQKLEKQNL-----EFNIAITGVNNA 173
 QY 140 -----FLGISIRPFANYVFLMYLFPRLRLRHFWTPKQQTDFLDIYHAF 186
 Db 174 GKSSLLNALKEDELGVSNIPETANLV--LSYGKSEAKI-YFMDKKEKMONILESSH- 229
 QY 187 RKQSHPEIISYLEKYI-----PLISDAGL-----RWRDLCTKIQGTH-- 226
 Db 230 -NADLKEFLDKDSVNIEDPIKDKPLIONIALCELNFSKAKNISALIKKIEIKSHLE 288
 QY 227 -----PAIHDLALRECFSNHPLG-----MNOQLAHVAKLSRAMLLISYL 267
 Db 289 FLKNNISIVDTPGLDGVVQREIVNMEYLRSDFLIHLMNASQSLTKD-----ADFL 341
 QY 268 PPLPLRRLKTHHTVYIHQDLKALAKLGIGOL--TAQEVKS----- 305
 Db 342 VHCILNSRLSKFLIVLTKAD-LLSKDLEEVIVYTKESLSKSLVLDENLVEKIDFLCVS 400
 QY 306 ----ACYLGLNSTH-----IGEDRCRTWL-----GEMLIQCSLKEAFSLILHNVY 349
 Db 401 AKMASDFYKGLAKESLOKSGMQEPENYLFNELYAGESKIALRAYKKELHLELNIL 458

RESULT 15

H71867
 methionine--tRNA ligase (EC 6.1.1.10) - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99
 C:Date: 12-Feb-1999 #sequence.revision 12-Feb-1999 #text_change 16-Jul-1999

C:Accession: H71867
 R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
 ; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
 Nature 397, 176-180, 1999
 A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
 A:Reference number: A71800; MUID:99120557

A:Accession: H71867
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-656 <ARN>
 A:Cross-references: GB:AE001525; GB:AE001439; NID:g4155533; PIDN:AAD06543.1; PID:g415
 A:Experimental source: strain J99
 C:Genetics:
 A:Gene: metG
 C:Superfamily: methionine--tRNA ligase
 C:Keywords: aminocacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 4.8%; Score 92; DB 2; Length 656;
 Best Local Similarity 17.7%; Pred. No. 4.8; Indels 64; Gaps 8;

Matches 49; Conservative 51; Mismatches 113;

QY 28 IQLGRSLAMGAPSSKHLSPKADVKNLSYVYTKKAINGYHRRFLGRHPPFYLYT 87
 Db 204 LSTITSEKGLPLPKKN-DPKHVVYVWDALLNVASALG----- 243
 QY 88 IFMKGLQMLADAKKARIKTMMKHNKTFHQLPYREMEHLRQFRODYTKCLFLGIISIP 147
 Db 244 -YLNGLDNKMAHFEERARI--VGKDLIRFAIYWP-----AFLMSLNL 284
 QY 148 PRANYLVFLMYLFPRLRLRHFWTP-----PKQQTDFLDIYHAFRKQSHPEIISYLEKY 201
 Db 285 LF-----KQLCVGWMTIEGVKSKSLGVNLAQKIAMERYGIEELRYFLRE 331
 QY 202 IPLISDAGLRWR--LTDLCTKIQGTHPAIHDLALRECFSNHPLGNNQALHVKALS 259
 Db 332 VFGDGDGFSKALVERINANLNDGLNRLLGMAKKFENYSLKSTKITAVYPELEK 391
 QY 260 AMLL-----SYLPPPLRRLKTHHTVYIHQDLKALAK 292
 Db 392 AHQILDNANSFVPMQLKALEELFNIDYDLNKLIAK 428

Search completed: December 20, 2001, 10:12:00
 Job time: 128 sec